EIN DIGITALES LEITUNGSMODELL ALS HILFSMITTEL ZUR SPRACHSYNTHESE

DISSERTATION

zur Erlangung des Grades eines
Doktor-Ingenieurs
der Abteilung für Elektrotechnik
an der Ruhr-Universität Bochum

von

JÜRGEN BRAAS

Ahlen/Westfalen

Bochum 1981

Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41544719_0007

EIN DIGITALES LEITUNGSMODELL ALS HILFSMITTEL ZUR SPRACHSYNTHESE

DISSERTATION

zur Erlangung des Grades eines
Doktor-Ingenieurs
der Abteilung für Elektrotechnik
an der Ruhr-Universität Bochum

von

JÜRGEN BRAAS

Ahlen/Westfalen

Bochum 1981

UNIV.
LEIDEN
BIBL.

Dissertation eingereicht am: 21.10.1980

Tag der mündlichen Prüfung: 08.01.1981

Referent: Prof. Dr.-Ing. J. Blauert

Korreferent: Prof. Dr. sc. techn. A. Fettweis

Vorwort

Die vorliegende Arbeit entstand am Lehrstuhl für Allgemeine Elektrotechnik und Akustik der Ruhr-Universität Bochum. Dem Leiter des Lehrstuhls, Herrn Prof. Dr.-Ing. Jens Blauert möchte ich herzlich für sein starkes Interesse und die intensive Förderung der Arbeit danken.

Meinen Kollegen Herrn Dipl.-Ing. U. Buschmann, Herrn Dr.-Ing. K. Meerkötter und Herrn Dr.-Ing. W. Wegener möchte ich für ihre ständige Diskussionsbereitschaft danken.

Herrn Prof. Dr. sc. techn. Alfred Fettweis, dessen Forschungsarbeit diese Arbeit erst ermöglichte, danke ich für die Übernahme des Korreferates.

Inhaltsverzeichnis

		Seite
1.	Einleitung	1
2.	Digitales Leitungsmodell	3
2.1	Elektrische Ersatzschaltung der akustischen Leitung	3
2.2	Einige Grundlagen der Wellendigitalfilter	5
2.3	Realisierung des Leitungsmodells	13
2.4	Übertragungseigenschaften des Leitungsmodells	16
3.	Ermittlung von Koeffizienten zur Sprach-erzeugung durch das Leitungsmodell	29
3.1	Methoden zur Bestimmung der Querschnitts-flächenfunktion des Vokaltrakts	30
3.1.1	Ermittlung der Querschnittsflächenfunktion aus akustischen Messungen	31
3.1.2	Ermittlung der Querschnittsflächenfunktion aus dem Sprachsignal	34
3.2	Lineare Prädiktion	36
3.2.1	Grundprinzip der linearen Prädiktion	37
3.2.2	Lösungsmethoden	43
3.2.3	Zusammenhang zwischen Prädiktor- und Adaptorkoeffizienten	46
3.2.4	Gewähltes Verfahren und praktische Gewinnung der Koeffizienten	53
3.2.5	Parametereinflüsse auf die lineare Prädiktion	56
3.2.5.1	Die Ordnung der Prädiktion	56
3.2.5.2	Einfluß der Fensterlänge und der Fensterform	62
3.2.5.3	Einfluß der Position des Fensters	66
4.	Anregungsfunktion	70
4.1	Anregungsfunktion für das Vokaltraktmodell	70
4.1.1	Stimmhafte Anregung	70
4.1.2	Stimmlose Anregung	72
4.2	Anregungsfunktion für das Prädiktionsfilter	72
4.2.1	Stimmhafte Anregung	72
4.2.2	Stimmlose Anregung	73
4.3	Realisierung des Anregungsgenerators	74
4.4	Gewinnung der Anregungsparameter aus natürlicher Sprache	75

4.4.1 Sprachgrundfrequenz 75
4.4.2 Amplitude 81

5. Rechnersimulation 83
5.1 Koeffizientenwortlänge 83
5.2 Aussteuerbarkeit 88
5.3 Maßnahmen zur Sicherung der Stabilität 91

6. Erste Synthesebeispiele 94

7. Zusammenfassung 97

Literaturverzeichnis 99

Anhang 107

Lebenslauf 115

1. Einleitung

Schon lange sind die Menschen von der Möglichkeit, Sprache auf andere als die natürliche Weise zu erzeugen, fasziniert (ausführlicher historischer Überblick z.B. in KÖSTER, 1973; FLANAGAN, 1972). Mit mechanischen Modellen wurden erste Erfolge Ende des 18. Jahrhunderts von KRATZENSTEIN und von VON KEMPELEN erzielt. Mit der Entwicklung der Elektrotechnik vergrößerte sich auch das Interesse an der synthetischen Spracherzeugung. Dabei können die entstandenen Sprachsynthesatoren in zwei verschiedene Klassen eingeteilt werden.

Die erste Klasse umfaßt die sogenannten "Terminal-Analog" Synthesatoren (FLANAGAN, 1957), die ohne die physiologischen und physikalischen Eigenschaften der natürlichen Entstehung von Sprache zu beachten das Ein-Ausgangsverhalten nachbilden. In der zweiten Klasse wird die Spracherzeugung durch die Nachbildung des Vokaltraktes, der Anregungsquelle Glottis und dem Abstrahlverhalten des Mundes erreicht. (Umfangreiche Literaturhinweise auf Spracherzeugungssysteme beider Klassen finden sich z.B. in FLANAGAN, 1972; FLANAGAN & RABINER, eds., 1973; FLANAGAN, 1976; RABINER & SCHAFER, 1976.)

In dieser Arbeit wird ein digitales Leitungsmodell vorgestellt, das in beiden Klassen der Spracherzeugungsmodelle eingesetzt werden kann. Ausgehend von der Beschreibung einer inhomogenen, verlustlosen Leitung durch die Aneinandereihung von Leitungselementen jeweils konstanten Querschnitts bietet sich eine digitale Realisierung entsprechend der Theorie der von FETTWEIS 1970 eingeführten Wellendigitalfilter an (BRAAS, 1977), da dort ein solches Bauelement vorhanden ist. Es wird gezeigt, daß dieses digitale Leitungsmodell auch als Prädiktionsfilter in einem Sprachsynthesesystem der ersten Klasse eingesetzt werden kann.

Da die Gewinnung von Parametern für ein System der zweiten Klasse, wie eine kritische Betrachtung der vorhandenen Verfahren zeigt, ohne unverhältnismäßig hohen experimentellen Aufwand z.Zt. mit genügender Genauigkeit nicht möglich ist,

wird mit dem digitalen Leitungsmodell ein Spracherzeugungssystem der "Terminal-Analog" Klasse aufgebaut. Um beim später geplanten Einsatz in einem System zur Sprachausgabe aus Text mit unbegrenztem Wortschatz Möglichkeiten der Optimierung von Parametern zu haben, wird auf möglichst hohe Flexibilität des aufgebauten Systems geachtet.

Zur intensiven Analyse vor dem schaltungsmäßigen Aufbau wird eine Simulation auf dem Laborrechner durchgeführt. Experimentell kann damit insbesondere nachgewiesen werden, daß nur eine von zwei Betriebsarten, die wie gezeigt wird, möglich sind, bei dem in Festkommaarithmetik arbeitenden Leitungsmodell sinnvoll ist. Für diese Beschränkung auf eine Betriebsart kann auch eine in der Theorie der Wellendigitalfilter beruhende Begründung gegeben werden.

Zur Gewinnung aller nötigen Parameter aus natürlicher Sprache für das sowohl schaltungsmäßig aufgebaute als auch als Simulationsprogramm vorhandene Spracherzeugungssystem werden geeignete Programme auf dem Laborrechner implementiert. Mit den damit gewonnenen Parametersätzen können erste Synthesebeispiele erzeugt werden.

2. Digitales Leitungsmodell

2.1 Elektrische Ersatzschaltung der akustischen Leitung

Schon seit geraumer Zeit wird als Modell für die Spracherzeugung die akustische Leitung untersucht. Dabei ist man insbesondere an einem elektrischen Äquivalent interessiert (DUNN, 1950; STEVENS et al., 1953; FANT, 1960). Man stellt dazu zunächst einmal den Vokal-Trakt in einer Flächenfunktion dar, die die Querschnittsflächen senkrecht zur Ausbreitungsrichtung des Luftstroms von der Glottis bis zu den Lippen darstellt. Diese Flächenfunktion wird nun durch eine genügend kleine Anzahl von gleichlangen Rohrstücken, die über dieser Länge konstante Querschnittsfläche haben, angenähert. Ein solches Element der Länge l kann durch das in Bild 2.1 dargestellte Zweitor dargestellt werden. Dabei entspricht dem Druck p im akustischen System die

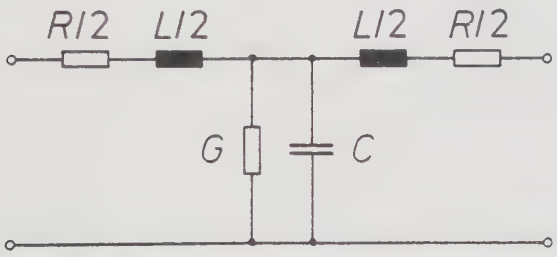


Bild 2.1 Elektrisches Ersatzschaltbild eines Rohrstücks mit konstanter Querschnittsfläche

Spannung u im elektrischen Modell und dem Schallfluß q der Strom i. Die Gültigkeit des elektrischen Modells ist auf den Bereich der Ausbreitung ebener Wellen beschränkt. Die Wellenlänge der sich ausbreitenden Welle sollte dazu größer als der Durchmesser sein. Die auf die Länge bezogenen Werte der Bauelemente des elektrischen Modells sind mit den physikalischen Größen des akustischen Rohres folgendermaßen verknüpft.

$$L = \frac{\rho}{A} \qquad C = \frac{A}{\rho c^2}$$

$$R = \frac{S}{A^2}\sqrt{\frac{\omega \rho \mu}{2}} \qquad G = S\,\frac{\eta - 1}{\rho c^2}\sqrt{\frac{\lambda_w \omega}{2 c_p \cdot \rho}} \tag{2.1}$$

Dabei ist A die Querschnittsfläche, S der Umfang des Rohres, ρ die Dichte der Luft, c die Schallgeschwindigkeit, μ der Viskositätskoeffizient, λ_w der Wärmeleitkoeffizient, η die Adiabaten Konstante und c_p die spezifische Wärmekapazität von Luft bei konstantem Druck (FLANAGAN, 1972). Das in Bild 2.1 dargestellte Ersatzschaltbild ist identisch mit dem einer elektrischen Leitung, deren Berechnungsverfahren sich somit hier anwenden lassen. Man erhält also den Leitungswellenwiderstand

$$Z_o = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \tag{2.2}$$

und die Ausbreitungskonstante

$$\gamma = \sqrt{(R + j\omega L)(G + j\omega C)} \tag{2.3}$$

Für den Fall der Verlustlosigkeit, den wir im folgenden betrachten wollen, wird der Leitungswellenwiderstand rein reell und ist nur noch von der Querschnittsfläche abhängig.

$$Z_o = \frac{\rho \cdot c}{A} \tag{2.4}$$

und die Ausbreitungskonstante reduziert sich auf

$$\gamma = j\,\frac{\omega}{c} \tag{2.5}$$

Ein solches homogenes, verlustloses Leitungsstück wird auch als "unit element" bezeichnet (FRAITURE & NEIRYNCK, 1969).

Bei der digitalen Nachbildung dieses Elements bietet sich na-

türlich an, auf eine Theorie zurückzugreifen, in der ein solches Element existiert. Dies ist bei der Theorie der Wellendigitalfilter der Fall.

2.2 Einige Grundlagen der Wellendigitalfilter

Neben den herkömmlichen Digitalfiltern (KAISER, 1966; IEEE Trans.AU-16, 1968; GOLD & RADER , 1969; SCHÜSSLER, 1973; OPPENHEIM & SCHAFER, 1975; RABINER & GOLD, 1975; HAMMING, 1977), die direkt aus der Übertragungsfunktion hergeleitet werden können und bei denen die Kaskadenfilter größere Bedeutung erlangt haben, wurden 1970 von Fettweis die Wellendigitalfilter eingeführt (FETTWEIS, 1970a; 1971a). Sie bieten eine Reihe von Vorteilen gegenüber den anderen Strukturen. Zunächst ist eine geringere Empfindlichkeit gegenüber Toleranzen der Koeffizienten zu nennen, die zunächst mit der Nachbildung der besonders toleranzunempfindlichen beidseitig resistiv abgeschlossenen, verlustfreien Abzweigfilter begründet wurde (FETTWEIS, 1960). Später wurde diese Eigenschaft mit Hilfe des neu eingeführten Begriffes der Pseudoleistung direkt bewiesen (FETTWEIS, 1972a).

Weiterhin läßt sich mit Hilfe des Pseudoleistungskonzeptes nachweisen, daß sich durch technisch einfach zu realisierende Maßnahmen parasitäre Schwingungen (Grenzzyklusschwingungen) in einem Wellendigitalfilter vermeiden lassen (FETTWEIS & MEERKÖTTER, 1974; 1975a). Diese Schwingungen sind das Ergebnis des instabilen Verhaltens, das durch die von der endlichen Signalwortlänge verursachten Nichtlinearität in jedem rekursiven Filter auftreten kann (KAISER, 1976; CLAASEN et al., 1976). Sie sind erkennbar z.B., wenn das Ausgangssignal des Digitalfilters nach dem Abschalten des Eingangssignals nicht wie erwünscht gegen Null konvergiert, sondern nichtabklingende Schwingungen ausführt. Für Wellendigitalfilter und nichtrekursive Digitalfilter wurde gezeigt, daß diese Stabilität auch bei Anordnung von vielen Filtern in einer Schleife eines Nachrichtenübertragungssystems erhalten bleibt (FETTWEIS & MEERKÖTTER, 1977). Weiterhin wurde gezeigt, daß mit Hilfe der Einführung des Begriffes der inkrementalen Pseudopassivität die Antwortstabilität garantiert werden kann (MEERKÖTTER, 1979). Dies ist die Forderung, daß Störungen des Ausgangssignals, verursacht

durch den Überlauf eines Addierers, wieder abklingen, vorausgesetzt, daß das Eingangssignal überlauffreien Betrieb ermöglicht.

Um die Möglichkeit zu haben, ein zeitkontinuierliches Filter durch ein Digitalfilter nachzubilden, werden folgende zwei Vereinbarungen getroffen:

1. Die komplexe Frequenz $p = \sigma + j\omega$ wird durch die Frequenzvariable

$$\psi = \tanh(pT/2) = (1-e^{-pT})/(1+e^{-pT}) \qquad (2.6)$$

ersetzt, wobei $T > 0$ die Betriebsperiode des Digitalfilters ist.

2. Als Größen in den Signalflußdiagrammen werden nicht die im kontinuierlichen Filter auftretenden Ströme und Spannungen benutzt, sondern im allgemeinen als Wellen bezeichnete Linearkombinationen von ihnen.

Wie aus (2.6) zu ersehen ist, ist die Frequenzvariable ψ über die bilineare Transformation

$$\psi = \frac{z - 1}{z + 1} \qquad (2.7)$$

mit der bei der Behandlung diskreter Signale üblichen Variablen $z = e^{pT}$ verknüpft. Da durch (2.6) der für Digitalfilter nutzbare Streifen der p-Ebene auf die ψ-Ebene abgebildet wird, läßt sich für alle Digitalfilter z.B. das Problem der Dämpfungsapproximation auf das entsprechende Problem für ein kontinuierliches Filter zurückführen (siehe z.B. SCHÜSSLER, 1973). Eine über diese nützliche Eigenschaft der Frequenzvariablen hinausgehende Bedeutung hat sie in der Theorie der Wellendigitalfilter. Da durch sie die Impedanzen der reaktiven Elemente eines LC-Filters definiert werden, ist eine Nachbildung dieser Elemente durch die Verzögerungen des Digitalfilter möglich (FETTWEIS, 1970a; 1971a).

Als Signalgrößen werden Wellen benutzt, da eine Verwendung der im LC-Filter auftretenden Spannungen und Ströme zu nichtrealisierbaren Signalflußdiagrammen führt (FETTWEIS, 1971a, 1976).

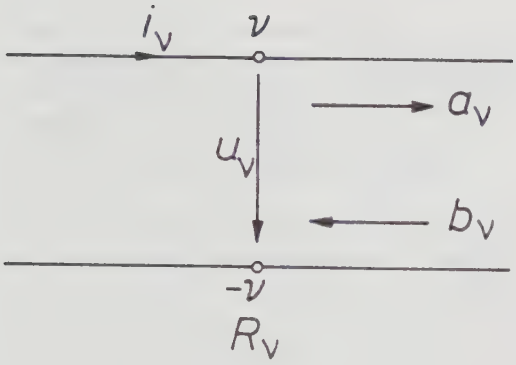


Bild 2.2
Definitionen der Wellen a_ν, b_ν am Tor ν

$$a_\nu = u_\nu + R_\nu \cdot i_\nu$$
$$b_\nu = u_\nu - R_\nu \cdot i_\nu$$

Um an einem Tor ν (Bild 2.2) die auftretenden Wellen zu definieren, ordnet man diesem Tor eine positive Konstante $R_\nu = 1/G_\nu$ zu, die die Dimension eines Widerstandes hat und Torwiderstand genannt wird. Als Spannungswelle wird dann die hinlaufende Welle a_ν und die rücklaufende Welle b_ν durch

$$a_\nu = u_\nu + R_\nu \cdot i_\nu \qquad (2.8a)$$

bzw.

$$b_\nu = u_\nu - R_\nu \cdot i_\nu \qquad (2.8b)$$

definiert. Dabei ist u_ν der Momentanwert der Spannung und i_ν der Momentanwert des Stromes. Allgemein können auch statt der Spannungswellen Stromwellen

$$a_\nu = i_\nu + G_\nu \cdot u_\nu \qquad (2.9a)$$

bzw.

$$b_\nu = i_\nu - G_\nu \cdot u_\nu \qquad (2.9b)$$

oder andere Linearkombinationen von u und i benutzt werden, was aber für unsere weiteren Betrachtungen keine Vorteile bringt, aber auch keine Einschränkung der Allgemeinheit bedeutet (FETTWEIS, 1973; MEERKÖTTER, 1979). Wir werden deshalb im weiteren nur Spannungswellen betrachten.

Um ein Signalflußdiagramm für ein Wellendigitalfilter zu er-

stellen, wird das nachzubildende (zeitkontinuierliche) Filter, das als Referenzfilter bezeichnet wird, in seine Elemente und Quellen sowie in ein Verbindungsnetzwerk zerlegt. Dieses Verbindungsnetzwerk verknüpft mit "idealen Drähten" die Bausteine des Filters und wird auch als Verbindungsmehrtor bezeichnet.

Für jeden Baustein wird ein Wellenflußdiagramm erstellt. Um eine möglichst einfache digitale Realisierung zu erreichen, wird der Torwiderstand entsprechend ausgewählt. So wird der Torwiderstand bei Bausteinen, die durch einen Parameter, der die Dimension eines Widerstandes hat, beschrieben werden, wie z.B. eine Kapazität, eine Induktivität gleich diesem Parameter gewählt. In der Tabelle 2.1 sind die wichtigsten nichtreaktiven Eintorelemente mit ihren Definitionsgleichungen und den daraus hergeleiteten Wellenflußdiagrammen dargestellt (MEERKÖTTER, 1979). Tabelle 2.2 zeigt die wichtigsten reaktiven Elemente und ihre Wellenflußdiagramme. Die in den Definitionsgleichungen auftretenden Größen U_1, U_2, I_1 und I_2 beschreiben die komplexen Amplituden der Signale u_1, u_2, i_1 bzw. i_2. Außer den beiden Elementen Kapazität und Induktivität, ist in dieser Tabelle der Sonderfall eines quasireziproken Leitungselements (Abk.: QUARL = quasi reciprocal line), das in der allgemeinsten Form durch die Gleichungen

$$b_1(t) = a_2(t - T/2 + \Delta)$$
$$b_2(t) = a_1(t - T/2 - \Delta) \qquad (2.10a, b)$$

beschrieben wird (FETTWEIS, 1970b; 1971b). Die Größe $T/2 > 0$ heißt mittlere Verzögerungszeit und die Größe Δ differentielle Verzögerungszeit. Falls Δ gleich Null ist, entspricht das quasireziproke Leitungselement dem schon vorher erwähnten gewöhnlichen (reziproken) Leitungselement ("unit element"). Es kann gezeigt werden (FETTWEIS, 1970b; 1971b; MEERKÖTTER, 1977), daß Veränderungen der differentiellen Verzögerungszeit der in einem Filter auftretenden quasireziproken Leitungselemente, lediglich eine Multiplikation der Übertragungsfunktion mit dem Faktor $e^{p\tau}$ (τ = reelle Konstante) bewirken. Da diese Multiplikation in den meisten Anwendungen irrelevant ist, kann man sich

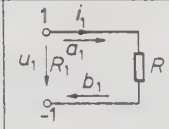
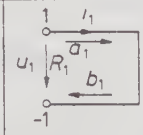
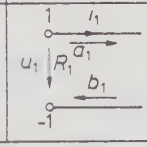
Element	Schaltbild	Spannungs-Strombeziehung	Wellen-beziehung	Wellenfluß-diagramm
Widerstand		$u_1 = R \cdot i_1$	$b_1 = 0$ $(R_1 = R)$	
Kurzschluß		$u_1 = 0$	$b_1 = -a_1$ (R_1 beliebig)	
Leerlauf		$i_1 = 0$	$b_1 = a_1$ (R_1 beliebig)	

Tabelle 2.1 Wellenflußdiagramme einiger nichtreaktiver Elemente

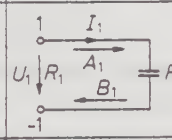
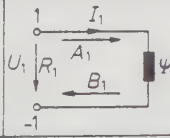
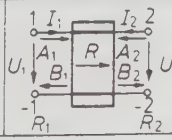
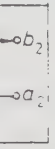
Element	Schaltbild	Spannungs-Strombeziehung	Wellen-beziehung	Wellenfluß-diagramm
Kapazitat		$U_1 = I_1 R/\Psi$	$b_1(t) = a_1(t-T)$ $(R_1 = R)$	
Induk-tivität		$U = \Psi R I$	$b_1(t) = -a_1(t-T)$ $(R_1 = R)$	
(quasi reziprokes) Leitungs-element		$\Psi R I_1 = U_1 - (1+\Psi) U_2$ $\Psi R I_2 = U_2 - (1-\Psi) U_1$	$b_1(t) = a_2(t)$ $b_2(t) = a_1(t-T)$ $(R_1 = R_2 = R)$	

Tabelle 2.2 Wellenflußdiagramme einiger reaktiver Elemente

bei Verwendung von quasireziproken Leitungselementen auf beliebige Sonderfälle beschränken. Insbesondere auf den in der Tabelle 2.2 angegebenen Fall $\Delta = T/2$ und den Fall $\Delta = -T/2$.

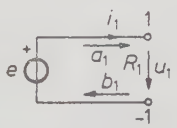
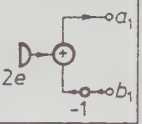
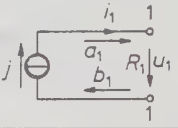
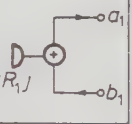
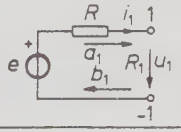
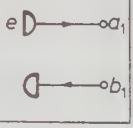
Quelle	Schaltbild	Spannungs-Strombeziehung	Wellen-beziehung	Wellenfluß-diagramm
Ideale Spannungsquelle		$u_1 = e$	$a_1 = 2e - b_1$ (R_1 beliebig)	
Ideale Stromquelle		$i_1 = j$	$a_1 = 2R_1 j + b_1$ (R_1 beliebig)	
Spannungsquelle mit Widerstand		$u_1 = e - R_1 i_1$	$a_1 = e$ ($R_1 = R$)	

Tabelle 2.3 Wellenflußdiagramme einiger Quellen

In Tabelle 2.3 sind die entsprechenden Größen und Diagramme der idealen Spannungs- und Stromquelle sowie der widerstandsbehafteten Spannungsquelle dargestellt (MEERKÖTTER, 1979). Die widerstandsbehaftete Stromquelle ist nicht aufgeführt, da ihr äußeres Verhalten identisch mit dem der Spannungsquelle ist.

Die digitalen Realisierungen der in den Tabellen dargestellten Wellenflußdiagramme werden als Wellendigitalbausteine bezeichnet, oder auch, wenn keine Verwechslung mit dem Referenzfilter zu befürchten ist, einfach als Bausteine. Sie tragen üblicherweise die Bezeichnung der entsprechenden Bausteine des Referenzfilters (MEERKÖTTER, 1979). Um aus den bisher dargestellten Bausteinen ein Wellendigitalfilter aufzubauen, benötigt man noch eine digitale Nachbildung, des bei der Zerlegung des entsprechenden Referenzfilters entstandenen Verbindungsnetzwerkes. Diese Verbindungsmehrtore werden durch als Adaptoren bezeichnete Mehrtor-Bausteine digital realisiert. Sie verbinden die Bausteine untereinander und passen die verschiedenen

Torwiderstände aneinander an. Die auftretenden Parallelverbindungen (Bild 2.3a) werden durch einen n-Tor-Paralleladaptor (Bild 2.3b), die auftretenden Serienverbindungen (Bild 2.4a)

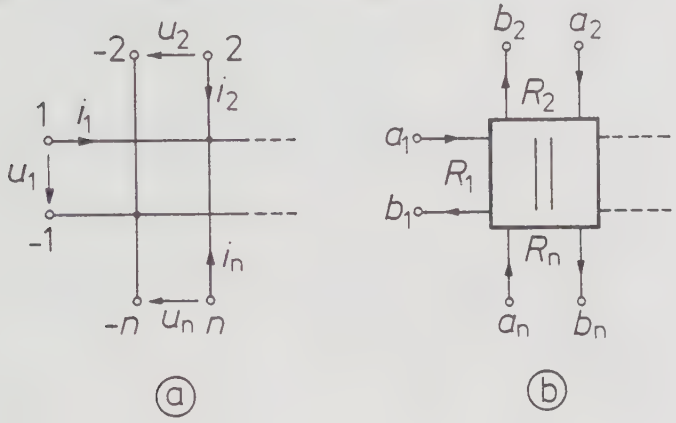


Bild 2.3 (a) Parallelverbindung von n Toren
(b) n-Tor-Paralleladaptor

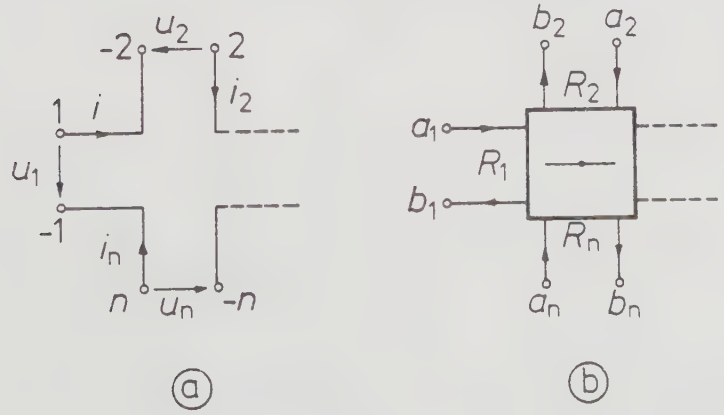


Bild 2.4 (a) Serienverbindung von n Toren
(b) n-Tor-Serienadaptor

durch einen n-Tor-Serienadaptor (Bild 2.4b) hergestellt (FETTWEIS, 1970a; FETTWEIS & MEERKÖTTER, 1975b). Auch bei der Verbindung von zwei Toren ist im allgemeinen ein Zweitor-Parallel-oder-Serienadaptor notwendig. Eine direkte Verbindung zweier Tore ohne den Einsatz eines Adaptors ist nur dann möglich, wenn beide Tore den gleichen Torwiderstand haben.

Für unseren Anwendungsfall der Hintereinanderschaltung von Leitungselementen verschiedener Leitungswellenwiderstände benötigen wir einen Zweitoradaptor, der die in Bild 2.5a dargestellte Verbindung zweier Tore mit den Torwiderstanden R_1 und R_2 herstellt. Mit den Bedingungen $i_1 = -i_2$ und $u_1 = u_2$ ergeben sich mit (2.8) für das zu realisierende Signalflußdiagramm folgende Wellenbeziehungen

$$b_1 = a_2 + \alpha\,(a_2 - a_1) \qquad (2.11a)$$

$$b_2 = a_1 + \alpha\,(a_2 - a_1) \qquad (2.11b)$$

mit $$\alpha = (R_1 - R_2)/(R_1 + R_2) \qquad (2.12)$$

Eine mögliche Realisierung dieser Gleichungen ist in Bild 2.5b gezeigt. Da eine solche Parallelverbindung auch als Serienverbindung aufgefaßt werden kann, stellt man den zugehörigen Adaptor wie in Bild 2.5c dar (FETTWEIS, 1971a).

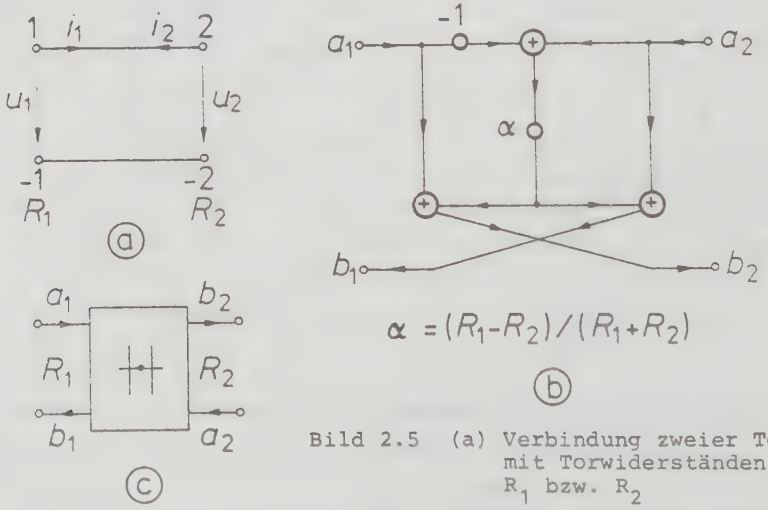


Bild 2.5 (a) Verbindung zweier Tore mit Torwiderständen R_1 bzw. R_2

(b) Eine mögliche Realisierung von (a)
(c) Blockdarstellung des Zweitoradaptors

2.3 Realisierung des Leitungsmodells

Mit den im vorhergehenden Abschnitt vorgestellten Elementen können wir nun ein digitales Modell der Aneinanderreihung von akustischen Leitungsstücken mit konstanter Querschnittsfläche bilden. Bild 2.6 zeigt das Wellenflußdiagramm einer solchen Schaltung mit N Leitungselementen. Dabei fehlt in der Abbildung der Anschluß der Leitung an eine Spannungsquelle mit Wi-

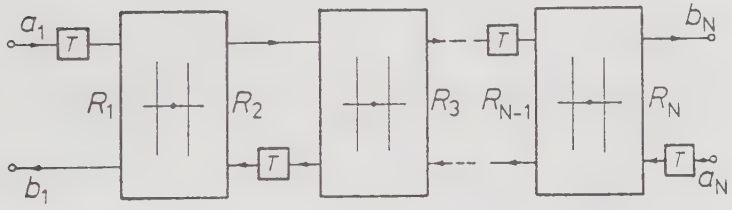


Bild 2.6 Wellenflußdiagramm eines Leitungsmodells mit N Leitungselementen

derstand und der Abschluß der Leitung mit einem Abstrahlwiderstand. Diese beiden Elemente könnten ebenfalls mit Hilfe des Zweitoradaptors mit dem ersten bzw. letzten Leitungselement verbunden werden. Die besonderen Eigenschaften dieser beiden Elemente erlauben es uns, eine einfachere Anbindung an das Leitungsmodell durchzuführen.

Wie aus Tabelle 2.1 hervorgeht, ist die reflektierte Welle an einem Widerstand immer identisch Null, falls der entsprechende Torwiderstand gleich dem Widerstand gewählt wird. Der in Bild 2.7a gezeigte Abschluß der Leitung mit einem Abstrahlwiderstand R_S läßt sich durch die Wellenbeziehungen

$$b_2 = a_1 - \alpha_S a_1 \qquad (2.13a)$$

$$b_1 = -\alpha_S a_1 \qquad (2.13b)$$

mit $\alpha_S = (R_N - R_S)/(R_N + R_S)$ beschreiben. Bis auf den bei b_2 auftretenden Faktor $1 - \alpha_S$ lassen sich die Wellenbeziehungen 2.13 durch das in Bild 2.7b dargestellte Flußdiagramm erfüllen.

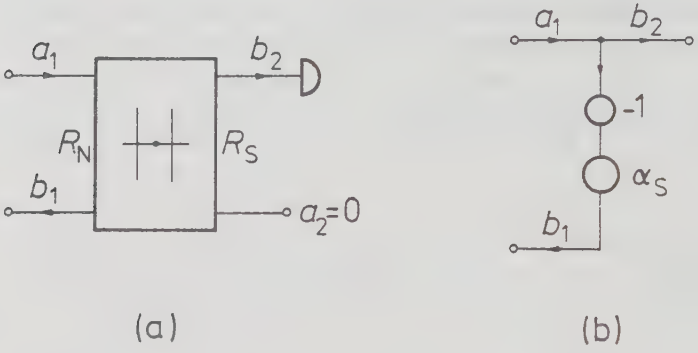


Bild 2.7 .(a) Abschluß der Leitung mit dem Abstrahlwiderstand R_S
(b) vereinfachte Realisierung von (a)

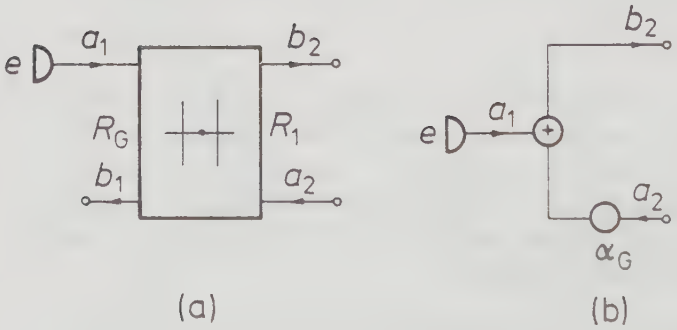


Bild 2.8 (a) Speisung der Leitung mit einer Spannungsquelle mit der Leerlaufspannung e und Innenwiderstand R_G

(b) vereinfachte Realisierung von (a)

Bild 2.8 zeigt das Wellenflußdiagramm des Anschlußes einer Spannungsquelle mit einem in Reihe liegenden Widerstand R_G an das erste Leitungselement. Die Wellenbeziehung

$$b_2 = (1 - \alpha_G)\, a_1 + \alpha_G\, a_2 \qquad (2.14)$$

mit $\alpha_G = (R_G - R_1)/(R_G + R_1)$, die diesen Anschluß beschreibt, läßt sich bis auf den Faktor $1 - \alpha_G$ bei a_1 durch das in Bild 2.8b gezeigte Wellenflußdiagramm realisieren.

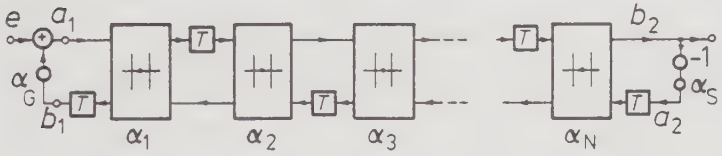


Bild 2.9 Wellenflußdiagramm des realisierten Leitungsmodells

Bild 2.9 zeigt das vollständige Wellenflußdiagramm des von BURGHARDT 1978 in einer Diplomarbeit aufgebauten Leitungsmodells. Es wurde nur ein Adaptor verwendet, der im Multiplexbetrieb die Berechnungen aller Adaptoren durchführt. Die Anzahl der Adaptoren kann dadurch zwischen N = 3 und N = 16 variiert werden. Außerdem läßt sich die Abtastfrequenz, d.h. die Frequenz mit der sich die Eingangs- bzw. Ausgangswerte ändern, zwischen 4 und 17 kHz einstellen. Die Signalwortlänge beträgt 16 bit, während die Wortlänge der Koeffizienten 8 bit beträgt. Alle auftretenden Zahlen werden als Festkommazahlen im Zweierkomplement dargestellt. Um bei der Realisierung der Koeffizienten α_G und α_S nicht jeweils einen Multiplizierer einsetzen zu müssen, wurde für den Koeffizienten α_G die Form $\alpha_G = \delta \cdot 2^{-1} + \delta \cdot 2^{-2} + \delta \cdot 2^{-3}$ mit $\delta = 0;1$ gewählt. Die Kombination des Inverters und des Koeffizient α_S kann die Werte

$-1 \cdot \alpha_S = -1 + 2^{-K}$ mit K = 0,1,2,3,4 annehmen. Diese Koeffizientenwerte lassen sich durch Verschieben und Addieren der Signale verwirklichen und ersparen somit den Einsatz eines Multiplizierers. Prinzipiell könnte diese Aufgabe der Anbindung an die Quelle und den Abschluß auch durch Erhöhung der Adaptoranzahl um zwei erreicht werden. Da aber bei den Hauptanwendungen die beiden Koeffizieten konstant sind, erübrigt sich in diesen Fällen die Übertragung der konstanten Koeffizienten vom Rechner zum Filter.

2.4 Übertragungseigenschaften des Leitungsmodells

Der Vektor der einfallenden Wellen $\underline{a} = (a_1, a_2, \ldots, a_n)^T$ wird mit dem Vektor der reflektierten Wellen $\underline{b} = (b_1, b_2, \ldots, b_n)^T$ * bei einem Mehrtor mit der Beziehung

$$\underline{b} = \underline{S}\,\underline{a} \qquad (2.15)$$

verknüpft. Die Matrix $\underline{S}$ wird als Streumatrix bezeichnet. Da die hier auftretenden Wellen Spannungswellen sind, unterscheidet sich die Matrix $\underline{S}$ von der üblicherweise mit Leistungswellen hergeleiteten Streumatrix $\underline{S}_L$. Es gilt aber, daß

$$\underline{S} = \underline{R}^{\frac{1}{2}}\,\underline{S}_L\,\underline{R}^{-\frac{1}{2}} \qquad (2.16)$$

mit $\underline{R} = \mathrm{diag}\,(R_1, R_2, \ldots, R_n)$ ($R_i \mathrel{\hat{=}}$ Torwiderstand) ist (FETTWEIS, 1972b).

Bei der Zusammenschaltung von Zweitoren verknüpft man zweckmäßigerweise die Wellen des Tores 1 mit denen des Tores 2

$$\begin{pmatrix} b_1 \\ a_1 \end{pmatrix} = \underline{T} \begin{pmatrix} a_2 \\ b_2 \end{pmatrix} \qquad (2.17)$$

* Das hochgestellte T kennzeichnet die Transponierung des jeweiligen Vektors.

Dabei ist $\underline{T}$ die Transfermatrix, die sich aus den Elementen der Streumatrix $\underline{S}$ des Zweitores als

$$\underline{T} = \frac{1}{S_{21}} \begin{pmatrix} -\det \underline{S} & S_{11} \\ -S_{22} & 1 \end{pmatrix} \tag{2.18}$$

ergibt.

Für ein klassisches verlustfreies Zweitor ist die kanonische Form der Transfermatrix, die die Leistungswellen verknüpft (FETTWEIS, 1969), durch

$$\underline{T}_L = \frac{1}{f} \begin{pmatrix} \sigma g_* & h \\ \sigma h_* & g \end{pmatrix} \tag{2.19}$$

gegeben.

Für die kanonischen Polynome f, g und h gilt

1. f, g und h sind reelle Polynome der komplexen Frequenz p.
2. g ist ein Hurwitzpolynom.
3. Der Koeffizient der höchsten Potenz von p im Polynom f ist auf + 1 normiert.
4. $\sigma = \pm 1$
5. f, g und h sind durch

$$g g_* = h h_* + f f_* \tag{2.20}$$

untereinander verknüpft.

Bei allen Funktionen bezeichnet der tiefgestellte Stern die Hurwitzkonjungierte der jeweiligen Funktion. Für reelle Funktionen, die wir hier betrachten, bedeutet es das Ersetzen von p durch - p. Bei verlustlosen, reziproken Zweitoren ist f entweder gerade oder ungerade und es gilt

$$\sigma = \frac{f_*}{f} = \begin{cases} +1 \text{ für gerades } f \\ -1 \text{ für ungerades } f \end{cases} \tag{2.21}$$

Mit (2.16) ergibt sich die für Spannungswellen definierte Transfermatrix als

$$\underline{T} = \sqrt{\frac{R_1}{R_2}}\ \underline{T}_L \tag{2.22}$$

Beide Matrizen unterscheiden sich also nur durch einen frequenzunabhängigen Faktor. Die oben angegebenen Eigenschaften der Transfermatrix klassischer verlustfreier Zweitore lassen sich also auf Wellendigitalfilter übertragen. Für die Polynome f, g und h der über Spannungswellen definierten Transfermatrix $\underline{T}$ wird nur die Beziehung (2.20) zu

$$gg_* = hh_* + \frac{R_1}{R_2}\, ff_* \tag{2.23}$$

wobei die Frequenzvariable p durch ψ (2.6) ersetzt wird.

Geht man nun mit Hilfe von (2.7) auf die bei diskreten Systemen übliche Beschreibung im z-Bereich über, ergeben sich Polynome in z. Mit Hilfe der gleichen Transformation (2.7) folgt allgemein

$$y_*(z) = y\left(\frac{1}{z}\right), \tag{2.24}$$

d.h. das Ergebnis ist kein Polynom in z mehr. Durch Multiplikation mit z^n, wobei n der Grad des Polynoms g ist, ergibt sich für die drei Polynome f, g und h im z-Bereich folgende Definition der Hurwitzkonjugierten (ULLRICH, 1976)

$$\begin{aligned} f_*(z) &= z^n\, f\left(\frac{1}{z}\right) \\ g_*(z) &= z^n\, g\left(\frac{1}{z}\right) \\ h_*(z) &= z^n\, h\left(\frac{1}{z}\right) \end{aligned} \tag{2.25}$$

Damit ergeben sich für die kanonischen Polynome der im z-Bereich über Spannungswellen hergeleiteten Transfermatrix

$$\underline{T} = \frac{1}{f} \begin{pmatrix} \sigma g_* & h \\ \sigma h_* & g \end{pmatrix} \tag{2.26}$$

folgende Bedingungen:

1) f, g, h sind reelle Polynome in z,

2) die Nullstellen des g-Polynoms liegen innerhalb des Einheitskreises,

3) der zur höchsten Potenz von z gehörende Koeffizient von f ist gleich + 1,

4) $\sigma = \pm 1$,

5) f, g und h sind durch

$$gg_* = hh_* + \frac{R_1}{R_2} ff_* \tag{2.27}$$

mit einander verknüpft. (Der tiefgestellte Stern kennzeichnet die mit (2.25) beschriebene Operation.) Bei verlustlosen, reziproken Zweitoren gilt weiterhin

$$\sigma = \frac{f_* (z)}{f (z)} = \begin{cases} + 1 \text{ für } f_* (z) = f (z) \\ - 1 \text{ für } f_* (z) = - f (z) \end{cases} \tag{2.28}$$

Die kanonische Form der Transfermatrizen der beiden Elemente, die wir beim Aufbau des Filters benutzen, sind für den Adaptor

$$\underline{T} = \begin{pmatrix} \frac{1}{1-\alpha} & \frac{-\alpha}{1-\alpha} \\ \frac{-\alpha}{1-\alpha} & \frac{1}{1-\alpha} \end{pmatrix} \tag{2.29}$$

und für das Leitungselement

$$\underline{T} = \frac{1}{z} \begin{pmatrix} 1 & 0 \\ 0 & z \end{pmatrix} \tag{2.30}$$

wobei das Leitungselement als QUARL realisiert ist, um Polynome in z zu erhalten. Für den in Bild 2.7b dargestellten Abschluß des Filters ergibt sich

$$\begin{pmatrix} a_1 \\ b_1 \end{pmatrix} = \begin{pmatrix} 1 \\ -\alpha_s \end{pmatrix} b_2 \tag{2.31}$$

als Verknüpfung der auftretenden Größen. Für die in Bild 2.8b dargestellte Realisierung des Eingangs ergibt sich die Beziehung

$$e = (1 \ , \ -\alpha_G) \begin{pmatrix} b_2 \\ a_2 \end{pmatrix} \tag{2.32}$$

Das in Bild 2.9 dargestellte Wellenflußdiagramm wird somit durch

$$e = (1 \ , \ -\alpha_G) \ \underline{T}_R \begin{pmatrix} 1 \\ -\alpha_s \end{pmatrix} b_2 \tag{2.33}$$

beschrieben, wobei

$$\underline{T}_R = \frac{1}{f_R} \begin{pmatrix} g_{R*} & h_R \\ h_{R*} & g_R \end{pmatrix} \tag{2.34}$$

die durch den Index R gekennzeichnete resultierende Matrix des Produktes einer entsprechenden Anzahl von Elementtransfermatrizen (2.29, 2.30) ist. Die Übertragungsfunktion zwischen Ein- und Ausgangsquelle ergibt sich aus (2.33) und (2.34) zu

$$H(z) = \frac{B_2}{E} = \frac{1}{T_{R11}^{-\alpha_G} T_{R21}^{-\alpha_S} T_{R12}^{+\alpha_G\alpha_S} T_{R22}} \qquad (2.35)$$

Wie gezeigt wurde (FRAITURE & NEIRYNCK, 1969) erfüllen auch die sich durch Multiplikation der kanonischen Transfermatrizen ergebenden Polynome die oben aufgezählten Bedingungen. Schreiben wir die Übertragungsfunktion (2.35) also als Funktion der durch den Index R gekennzeichneten resultierenden Polynome, ergibt sich

$$H(z) = \frac{B_2}{E} = \frac{f_R}{q_{R_*}^{-\alpha_G} h_{R_*}^{-\alpha_S} h_R^{+\alpha_G\alpha_S} \sigma_R} \qquad (2.36)$$

Da nur die Transfermatrizen des Zweitoradaptors (2.29) und des Leitungselementes (2.30) miteinander multipliziert werden, können wir folgende Aussage über die Übertragungsfunktion H (z) (2.36) machen. Da

$$f_R(z) = z^N \quad (N \hat{=} \text{Anzahl der Leitungselemente}) \qquad (2.37)$$

liegen die Nullstellen der Übertragungsfunktion alle bei z = 0. Der Dämpfungsverlauf wird also nur durch die Pole bestimmt. Solche Filter werden in der Literatur häufig als "Allpolfilter" bezeichnet, wobei dann fälschlicherweise behauptet wird, die Übertragungsfunktion hätte keine Nullstellen (z.B. RABINER & SCHAFER, 1978).

Im folgenden zeigen wir, daß bei bekannter Übertragungsfunktion H (z) die Werte der charakteristischen Widerstände der aneinandergereihten Leitungselemente nicht eindeutig berechnet werden können. Gegeben sei eine Schaltung entsprechend Bild 2.10. Die Übertragungsfunktion zwischen Ausgangs- und Eingangwelle $H(z) = \frac{B_2}{A_1}$ dieser Schaltung ergibt durch Multiplikation mit $(1 - \alpha_S)(1 - \alpha_G)$ die Übertragungsfunktion (2.35) der realisierten Schaltung (Bild 2.9), so daß alle folgenden Aussagen auch für die realisierte Schaltung zutreffen. Die Übertragungsfunktion der Schaltung in Bild 2.10 ist

$$H(z) = \frac{B_2}{A_1} = \frac{-2R_s \cdot I_2}{E} \tag{2.38}$$

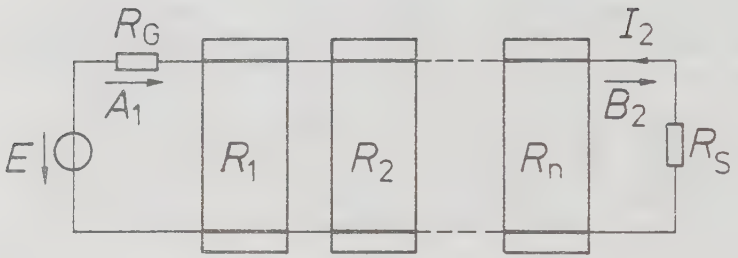


Bild 2.10 Inhomogene Leitung aus n Leitungselementen mit Speisung durch eine Spannungsquelle mit Innenwiderstand R_G und Abschluß durch Widerstand R_S

Ändert man nun die Reihenfolge der Leitungselemente und vertauscht R_G und R_S, erhält man die in Bild 2.11 dargestellte Schaltung. Die Übertragungsfunktion

$$H'(z) = \frac{B_1'}{A_2'} = \frac{-2R_G\, I_1'}{E} \tag{2.39}$$

läßt sich, da wegen der Reziprozität der Schaltung

$$I_1' = I_2 \tag{2.40}$$

ist, mit

$$H'(z) = \frac{R_G}{R_S} H(z) \tag{2.41}$$

aus der ursprünglichen Übertragungsfunktion (2.38) ermitteln. Die duale zu der in Bild 2.11 dargestellten Schaltung zeigt

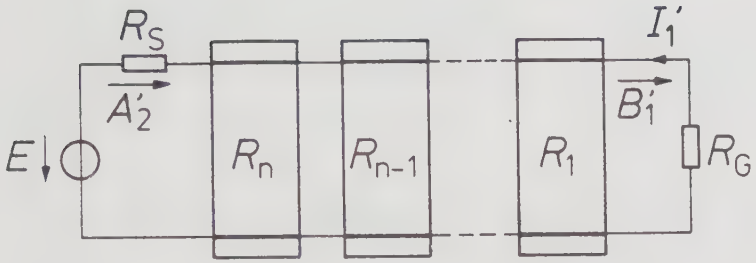


Bild 2.11 Inverser Betrieb der Schaltung nach Bild 2.10

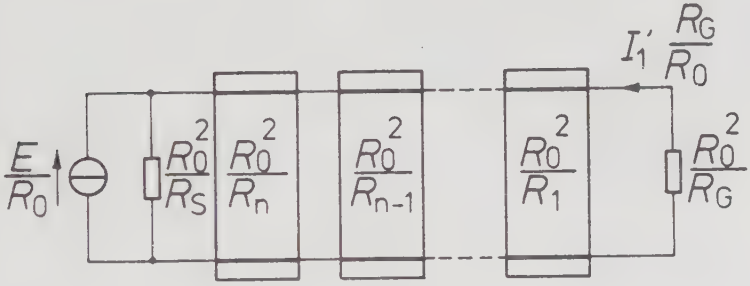


Bild 2.12 Zur Schaltung nach Bild 2.11 duale Schaltung

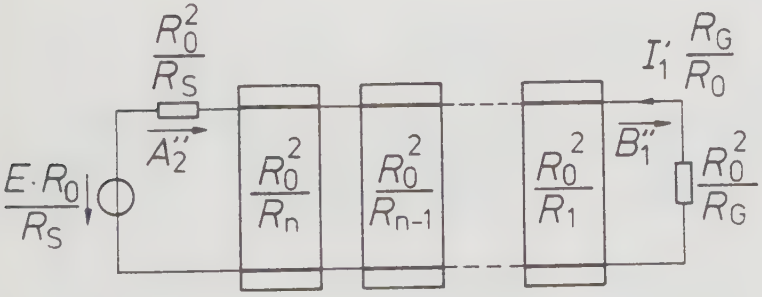


Bild 2.13 Schaltung mit äquivalentem Übertragungsverhalten wie Schaltung nach Bild 2.10

Bild 2.12. Wandelt man noch die entstandene Stromquelle in eine Spannungsquelle, so ergibt sich eine Schaltung entsprechend Bild 2.13. Die Übertragungsfunktion ergibt sich hierbei zu

$$H''(z) = \frac{B_1''}{A_2''} = \frac{-2\,\frac{R_o^2}{R_G}\cdot\frac{R_G}{R_o}\,I_1'}{E\,\frac{R_o}{R_S}} = \frac{-2\,R_S\,I_1'}{E} \qquad (2.42)$$

und ist somit wegen (2.40) identisch mit der Übertragungsfunktion (2.37) des ursprünglichen Netzwerkes. Für das realisierte Filter bedeutet diese Aussage, daß es mit dem Koeffizientensatz $\{\alpha_G, \alpha_1, \alpha_2, \ldots, \alpha_N, \alpha_S\}$ die gleiche Übertragungsfunktion hat wie mit dem Koeffizientensatz $\{\alpha_S, \alpha_N, \alpha_{N-1}, \ldots, \alpha_1, \alpha_G\}$. Das oben Gezeigte gilt zunächst nur für Werte von α_G und α_S ungleich Null bzw. Eins. Aber auch bei den beiden Sonderfällen $\alpha_G = 0$, $\alpha_S = 1$ - d.h. Spannungsquelle mit Widerstand am Eingang und Kurzschluß am Ausgang - sowie $\alpha_G = \alpha_S = 1$ - d.h. ideale Stromquelle am Eingang und Kurzschluß am Ausgang - können wir zeigen, daß das oben Gesagte gilt. Bild 2.14 verdeutlicht den Beweis für den ersten Sonderfall. Die Übertragungsfunktion zwischen Ausgangs- und Eingangswelle der Schaltung in Bild 2.14a ist

$$H(z) = \frac{B_2}{A_1} = \frac{-R_n \cdot I_2}{E} \qquad (2.43)$$

Für die Schaltung b gilt wegen der Reziprozität auch die Aussage der Gleichung (2.40) und für die zu ihr duale Schaltung - Teil c - ergibt sich als Übertragungsfunktion

$$H'(z) = \frac{B_1'}{A_2'} = \frac{-2\,\frac{R_o^2}{R_1}\cdot\frac{R_1}{R_o}\,I_1'}{2\,\frac{E}{R_o}\cdot\frac{R_o^2}{R_n}} = \frac{-R_n\,I_1'}{E} \qquad (2.44)$$

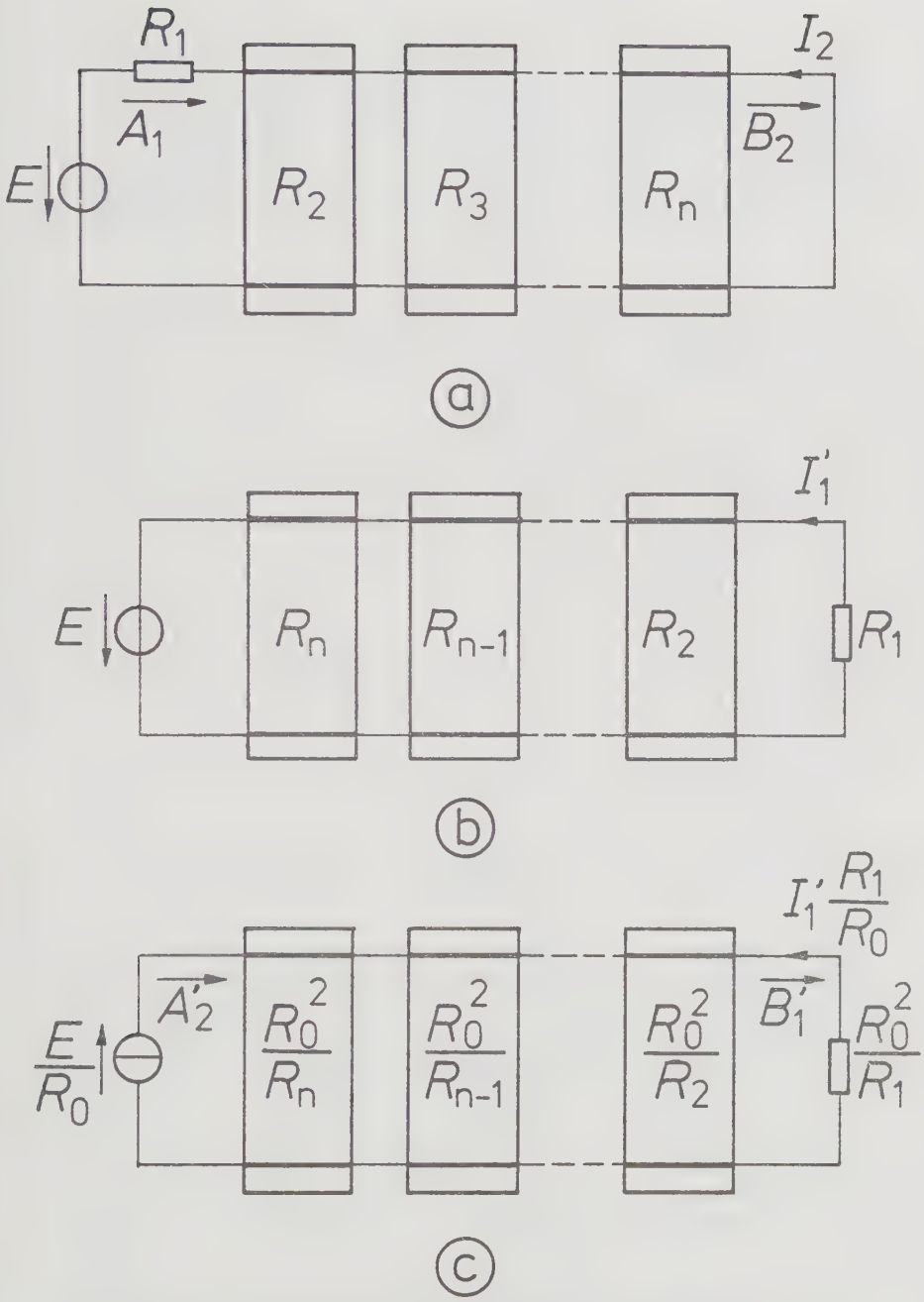


Bild 2.14 Veranschaulichung des Nachweises gleichen Übertragungsverhaltens von Schaltung (a) und (c)

Wegen (2.40) ist diese Übertragungsfunktion identisch mit der Übertragungsfunktion der Ausgangsschaltung - Teil a.

Der zweite Sonderfall ist in Bild 2.15 dargestellt. Teil a zeigt die Ausgangsschaltung, deren Übertragungsfunktion zwischen Ein- und Ausgangsstrom

$$G(z) = \frac{I_2}{I_1} = \frac{-\frac{B_2}{R_n}}{\frac{A_1}{R_1}} \tag{2.45}$$

ist. Die Übertragungsfunktion zwischen Ein- und Ausgangswelle ergibt sich somit zu

$$H(z) = \frac{B_2}{A_1} = -\frac{R_n}{R_1} G(z) \tag{2.46}$$

Auf Grund der Reziprozität des Netzwerkes gilt für die Schaltung in Teil b

$$\frac{U_2}{E} = G(z) \tag{2.47}$$

Bei der dazu dualen Schaltung - Teil c - gilt demnach

$$\frac{I_2'}{I_1'} = G(z) = \frac{-B_1' \Big/ \frac{R_o^2}{R_1}}{A_2' \Big/ \frac{R_o^2}{R_n}} \tag{2.48}$$

und somit auch

$$H'(z) = \frac{B_1'}{A_2'} = -\frac{R_n}{R_1} G(z) \tag{2.49}$$

Wird die Stromquelle, wie schon in Teil c von Bild 2.15 dar-

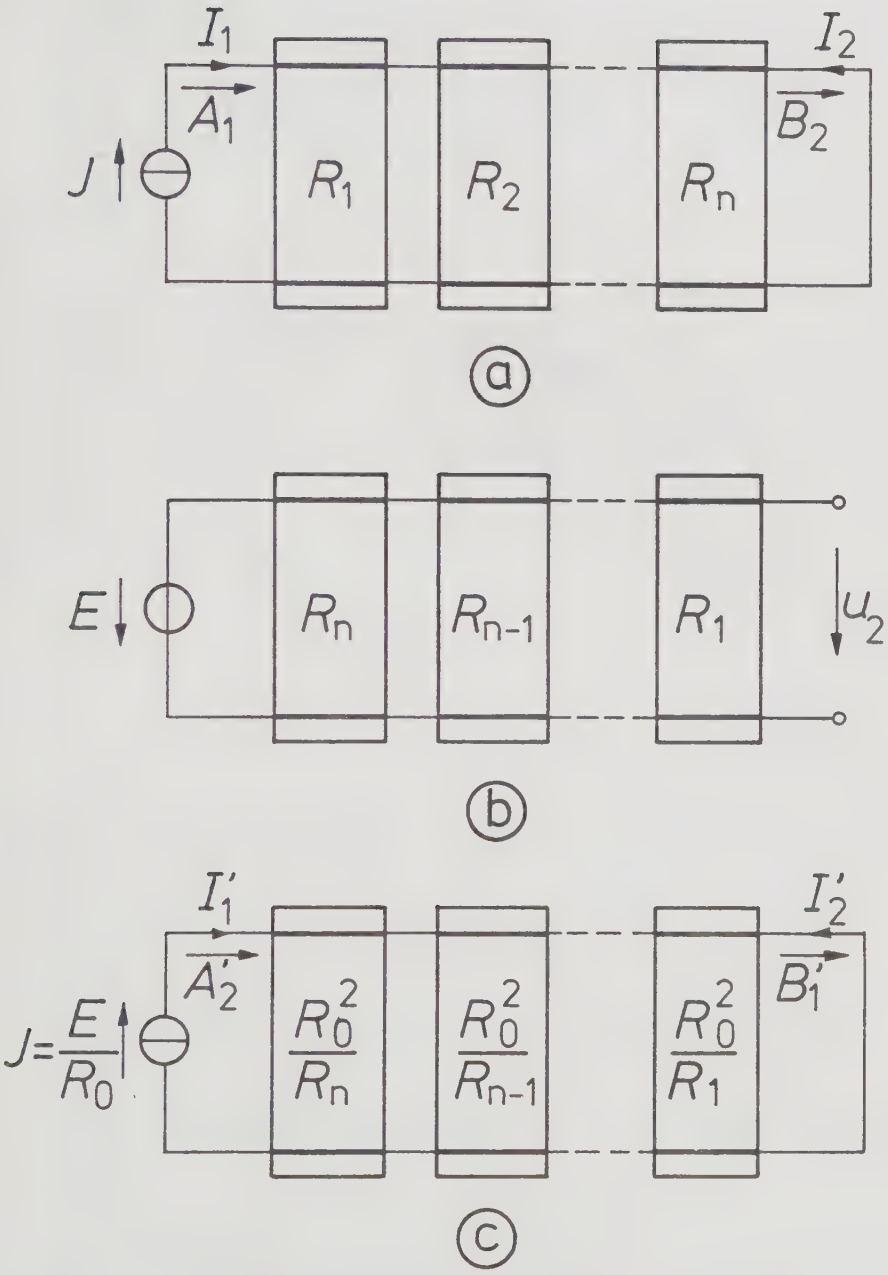


Bild 2.15 Veranschaulichung des Nachweises gleichen Übertragungsverhaltens von Schaltung (a) und (c)

gestellt, gleich der ursprünglichen Stromquelle gewählt, was möglich ist, da die Wahl von E und R_o beliebig ist, so wird auch die Übertragungsfunktion (2.49) der Schaltung in Teil c identisch mit der der Schaltung (2.46) nach Bild 2.15a.

3. Ermittlung von Koeffizienten zur Spracherzeugung durch das Leitungsmodell

Zur Spracherzeugung kann das im Kapitel 2 beschriebene digitale Leitungsmodell in zwei recht unterschiedlichen Modellen der menschlichen Spracherzeugung eingesetzt werden. Bild 3.1

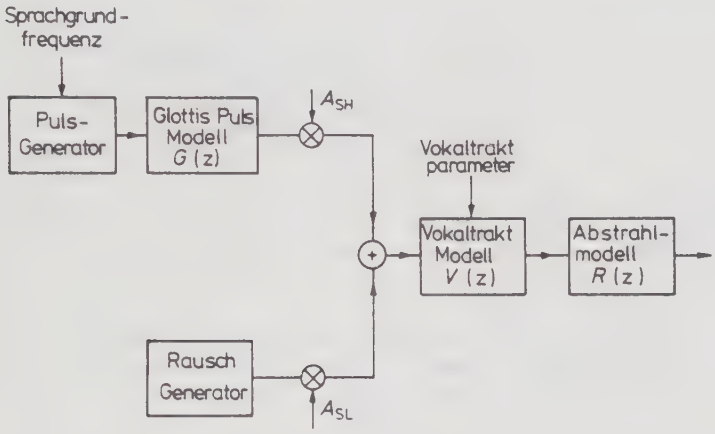


Bild 3.1 Digitales Spracherzeugungsmodell mit Berücksichtigung der physiologischen und physikalischen Gegebenheiten

zeigt ein digitales Modell, das versucht, die physiologischen und physikalischen Gegebenheiten zu berücksichtigen (RABINER & SCHAFER, 1978), wobei die Komponenten Anregung, Filterung im Vokaltrakt und Abstrahlung getrennt betrachtet werden. Faßt man diese drei Einzelsysteme zusammen, so ergibt sich das in Bild 3.2 dargestellte vereinfachte Modell zur Spracherzeugung. Das digitale Leitungsmodell kann, wie wir im folgenden sehen werden, in beiden Modellen eingesetzt werden. Dabei sind die Koeffizientensätze α, die jeweils das Übertragungsverhalten beschreiben, natürlich nicht gleich. In diesem Kapitel wollen wir Verfahren zur Ermittlung dieser Koeffizientensätze vorstellen.

Die Faktoren A_{SH} und A_{SL} werden im Kapitel 4.3 näher erläutert.

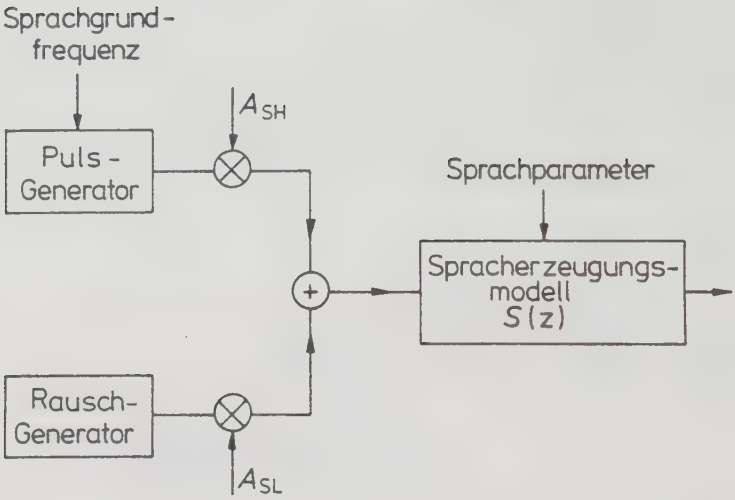


Bild 3.2 Digitales Spracherzeugungsmodell der "Terminal Analog" Klasse

3.1 Methoden zur Bestimmung der Querschnittsflächenfunktion des Vokaltraktes

Im ersten im Bild 3.1 dargestellten Modell wird das Vokaltrakt-Modell mit der Übertragungsfunktion V(z) durch das digitale Leitungsmodell beschrieben. Mit (2.4) und (2.12) ergeben sich die Adaptorkoeffizienten als

$$\alpha_i = - r_i \qquad (3.1)$$

aus den Reflektionskoeffizienten

$$r_i = \frac{A_i - A_{i+1}}{A_i + A_{i+1}} \quad , \qquad (3.2)$$

die zwischen den gleichlangen Rohrstücken, die den Vokaltrakt approximieren, auftreten. Mit der Funktion A(x), die die Querschnittsfläche über der Länge des Vokaltraktes darstellt, können wir diese Koeffizienten gewinnen. Eine ausführliche Dar-

stellung der Querschnittsflächenfunktionen aller Laute, die auf Röntgenaufnahmen beruht, liefert FANT, 1960. Um die Flächenfunktion auch als Funktion der Zeit darstellen zu können, werden ebenfalls Röntgenaufnahmen gemacht (z.B. HEINZ & STEVENS, 1964; 1965). Die dabei auftretenden Probleme der zweidimensionalen Darstellung des dreidimensionalen Problems werden sicherlich mit Hilfe der Computertomographie überwunden werden können, zur Zeit stehen aber noch keine ausreichenden Ergebnisse zur Verfügung (HEIKE, 1979). Um die Nachteile der Röntgenuntersuchungen zu vermeiden, sucht man andere Möglichkeiten, um den Querschnittsflächenverlauf zu ermitteln.

3.1.1 Ermittlung der Querschnittsflächenfunktion aus akustischen Messungen

Aufgrund der im Kapitel 2.1 vorgestellten elektrischen Ersatzschaltung für die akustische Leitung ist das Problem der Bestimmung der Querschnittsflächenfunktion identisch mit der Bestimmung des Leitungswellenwiderstandes als Funktion des Ortes bei einer inhomogenen, verlustlosen Leitung. Gegeben sei also eine solche Leitung, wobei man den elektrischen Ort y als Funktion des physikalischen Ortes x auf der Leitung als

$$y(x) = \int_0^x \{L(\eta)\ C\ (\eta)\}^{1/2}\, d\eta \qquad (3.3)$$

und die örtliche charakteristische Impedanz als

$$Z_o(y) = \{L[x(y)]/C[x(y)]\}^{1/2} \qquad (3.4)$$

definiert. Dabei sind L(x) und C(x) der auf die Länge bezogene Induktivitäts- bzw. Kapazitätsbelag. Die elektrische Länge Δ ergibt sich dann mit (3.3) und der Länge L_1 als

$$\Delta = \int_0^{L_1} \{L(\eta)\ C\ (\eta)\}^{1/2}\, d\eta \qquad (3.5)$$

Sind nun alle Pole und Nullstellen der Eingangsimpedanz Z_{ein} einer solchen Leitung bekannter Länge Δ und der Grenzwert

$$\lim_{|p|\to\infty} Z_{ein}(p)$$
$$(\mathrm{Re}\ p \geq 0)$$

bei Abschluß der Leitung mit einem Kurzschluß oder Leerlauf bekannt, so läßt sich $Z_o(y)$ eindeutig bestimmen (WOHLERS, 1969). Auf dieser Tatsache beruhen die von SCHROEDER, 1967 und MERMELSTEIN, 1967 entwickelten Verfahren. Da aber eine Ausbreitung von ebenen Wellen bei den Abmessungen des Vokaltraktes nur bis etwa 4 kHz gewährleistet ist, sind Pole und Nullstellen der gemessenen Eingangsimpedanz oberhalb dieser Frequenz bedeutungslos. MERMELSTEIN ermittelt zwar aus den ersten sechs Pol- und Nullstellen der Eingangsimpedanz die ersten sechs Koeffizienten eines in der Form

$$\log A(x) = \log A_o + \sum_{j=1}^{\infty} a_j \cos \frac{j\pi x}{L_1} \qquad (3.6)$$

angesetzten Logarithmus der Querschnittsfunktion, kann aber nicht beweisen, daß die Lösung eindeutig ist. Empirische Untersuchungen mit gegebenen Sätzen von Pol- und Nullstellen führten, wie MERMELSTEIN angibt, immer zu eindeutigen Lösungen.

Approximieren wir nun aber die Querschnittsflächenfunktion durch Aneinanderreihung von gleichlangen Rohrstücken konstanter Querschnittsfläche, so liegt durch die maximale Frequenz, bei der das elektrische Modell noch Gültigkeit hat, die Betriebsperiode T des Digitalfilters fest. Bei der oben angegebenen Frequenz von 4 kHz ergibt sich T = 125 µs; damit liegt auch die Laufzeit durch das Leitungsstück fest und somit bei einer Schallgeschwindigkeit von c = 340 m/s auch die Länge l = 2,125 cm. Bei einer Gesamtlänge L_1 = 17 cm des Vokaltraktes ergeben sich also N = 8 Rohrstücke zur Approximierung der Querschnittsflächenfunktion. Nach RICHARDS 1948 läßt sich nun jede Reaktanz der Ordnung n durch eine kurzgeschlossene oder leerlaufende Hintereinanderschaltung von n Leitungselementen realisieren. Ausgehend von der bekannten Eingangsreaktanz Z_{ein} lassen sich dann, wie Bild 3.3 verdeutlicht, rekursiv durch

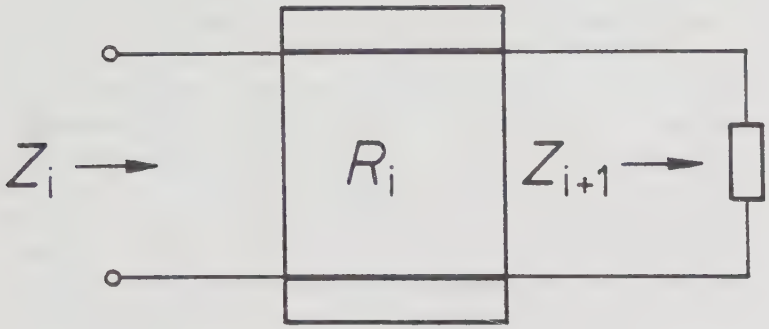


Bild 3.3 Veranschaulichung der Reaktanzrealisierung mit (3.7) nach RICHARDS, 1948

$$Z_{i+1} = R_i \frac{Z_i - \psi R_i}{R_i - \psi Z_i} \qquad (3.7)$$

mit $i = 1,2,\ldots,n$ und $Z_1 = Z_{ein}$ die charakteristischen Widerstände

$$R_i = Z_i \ (1) \qquad (3.8)$$

der einzelnen Leitungselemente berechnen. Die Kenntnis dieser Tatsache erlaubt es uns nun, in dem von SCHROEDER beschriebenen Verfahren die oben genannte Ermittlung der ersten sechs Fourierkoeffizienten des Logarithmus der Querschnittsflächenfunktion, durch die rekursive Berechnung der Querschnittsfläche der 8 gleichlangen Rohrstücke zu ersetzen. Die zur Berechnung benutzte Eingangsreaktanz ist dabei die z.B. mit einem quadratischen Fehlerkriterium bewertete beste Approximation der gemessenen Eingangsimpedanz. Damit haben wir dann im Gegensatz zu SCHROEDER UND MERMELSTEIN einen bewiesenen eindeutigen Zusammenhang zwischen der Eingangsreaktanz und den berechneten Querschnittsflächen. Um diese Methode der Querschnittsflächenbestimmung anwenden zu können, muß die Länge

des Vokaltraktes bekannt sein, und es muß sichergestellt sein, daß er an der Glottis schallhart, im elektrischen Modell also mit einem unendlich großen Widerstand abgeschlossen ist. Diese beiden Voraussetzungen sind in der Praxis nur unzureichend gegeben, so daß die Ermittlung der tatsächlichen Querschnittsflächenfunktion mit dieser Methode nicht sehr zuverlässig ist.

Ein von SONDHI & GOPINATH, 1971 entwickeltes Verfahren, das auch auf eine akustische Messung zurückgreift, benötigt weder die Information über die genaue Länge des Vokaltraktes noch die Kenntnis der Abschlußimpedanz an der Glottis. Bei diesem Verfahren wird die Impulsantwort an den Lippen gemessen, und zwar der Druck p als Antwort auf einen Dirac-Puls des Volumenflusses q. Es ist leicht einsichtig, daß die Impulsantwort $h(t)$ für $t \leq T$ unabhängig ist vom Querschnitt $A(x)$ des Vokaltraktes an Orten $x > c\ T/2 = L$, wobei c die Schallgeschwindigkeit ist. SONDHI und GOPINATH beweisen sowohl für den akustischen als auch für den äquivalenten Fall der inhomogenen, verlustlosen Leitung (GOPINATH & SONDHI, 1971), daß die Kenntnis der zeitlich begrenzten Impulsantwort ausreicht, um eindeutig den Querschnittsverlauf $A(x)$ bzw. $Z_o(y)$ (3.4) zu berechnen. Außer an einem Metallrohr und mit Rechnersimulation (SONDHI, 1977) haben sie die praktische Verwendbarkeit ihres Verfahrens nicht gezeigt. An anderer Stelle (DESCOUT et al., 1976; TOUSIGNANT et al., 1979) ist das Verfahren auch bei Messungen am Menschen benutzt worden. Aufgrund der meßtechnischen Schwierigkeiten war es aber bis jetzt nur möglich, eine relativ beschränkte Anzahl von Querschnittsflächenfunktionen zu ermitteln.

3.1.2 Ermittlung der Querschnittsflächenfunktion aus dem Sprachsignal

Um den großen meßtechnischen Aufwand der in Abschnitt 3.1.1 beschriebenen Verfahren zu vermeiden, wurde nach Methoden gesucht, die Querschnittsflächenfunktion aus dem Sprachsignal zu ermitteln. Aus dem in Bild 3.1 gezeigten Modell sieht man, daß dazu zunächst die Einflüsse der Abstrahlung und der Formung des Glottissignals eliminiert werden müssen. WAKITA, 1973 nähert diese Einflüsse durch einen 6 dB/Oktave Anstieg bzw.

einen 12 dB/Oktave Abfall an. Durch Filterung mit einem Hochpaß kompensiert er den resultierenden 6 dB/Oktave Abfall. Das so geglättete Spektrum betrachtet er als das der Übertragungsfunktion des verlustlosen Vokaltraktes mit der Abstrahlimpedanz Null und der Speisung durch eine Quelle mit Innenwiderstand. Die von WAKITA aufgestellte Behauptung, daß mit dieser Anhebung die richtige Querschnittsfunktion ermittelt wird, während sich ohne die Anhebung eine völlig andere, falsche ergibt, wird von SONDHI, 1977 und dem Verfasser nicht geteilt. Das einzige, was aus der Unterschiedlichkeit der ermittelten Querschnittsfunktionen geschlossen werden kann, ist, daß Annahmen über die Abstrahlung und die Glottissignalformung grossen Einfluß auf die ermittelte Querschnittsfunktion haben.

Auch ein von ATAL, 1970 vorgestelltes Verfahren, das von anderen Annahmen ausgeht, aber den gleichen Algorithmus wie WAKITA benutzt, wird später von diesem selber als nicht erfolgreich bezeichnet (ATAL, 1974). Auch mit komplizierteren Modellen und iterativer Anpassung der akustischen Größen des Modelles an das Sprachsignal (STRUBE, 1974) ergeben sich noch so viele Schwierigkeiten, daß auch solche Verfahren nicht erfolgversprechend erscheinen. Grundsätzlich können auch bei diesen Verfahren nur Sprachsignaldaten bis zu der bereits erwähnten Grenzfrequenz von 4 kHz ausgewertet werden, was die räumliche Auflösung begrenzt. Bei höherer örtlicher Auflösung, kleinerer Leitungslänge, ergeben sich dann im betrachteten Frequenzbereich bis 4 kHz Mehrdeutigkeiten, da völlig verschiedene Querschnittsverläufe, die aber physiologisch möglich sind, ununterscheidbare akustische Signale erzeugen (ATAL et al., 1978). Diese Erkenntnis erhärtet die Hypothese der kompensatorischen Artikulation, d.h. zwei Personen können den gleichen Laut mit unterschiedlichen Querschnittsflächenfunktionen des Vokaltraktes erzeugen. Außerdem gibt sie eine Erklärung für die Fähigkeit des Bauchredens, d.h. die Erzeugung gleicher Laute mit unterschiedlicher Mundöffnung.

Zusammenfassend kann also gesagt werden, daß z.Zt. einigermaßen verläßliche Ermittlungen der tatsächlichen Querschnittsflächenfunktion nur mit dem von SONDHI und GOPINATH, 1971 entwickelten

Impulsverfahren möglich sind, denn die Annahmen, die Voraussetzung für die Anwendbarkeit der Verfahren von ATAL und WAKITA sind, sind nach wie vor sehr unsicher (WAKITA, 1979). Es ist geplant, das Impulsverfahren, das schon bei der Messung der Trommelfellimpedanz unter Berücksichtigung der Querschnittsfunktion des Gehörgangs (JOSWIG, 1979) benutzt wurde, auch zur Ermittlung der Querschnittsfunktion des Vokaltraktes einzusetzen.

3.2 Lineare Prädiktion

Auf das in Bild 3.2 dargestellte Spracherzeugungsmodell lassen sich zur Bestimmung der Koeffizienten mathematische Methoden der Systemidentifikation anwenden. Setzt man den Zusammenhang zwischen Ausgangs- und Eingangsfolge des diskreten Systems als

$$y(n) = -\sum_{k=1}^{p} \beta_k \, y(n-k) + G \sum_{i=0}^{q} \gamma_i \, x(n-i) \qquad (3.9)$$

mit $\gamma_0 = 1$ an, so sieht man, daß die Ausgangswerte $y(n)$ Linearkombinationen zurückliegender Ausgangswerte und aktueller und zurückliegender Eingangswerte sind, und somit linear prädizierbar sind. Für dies Verfahren der Ermittlung der Koeffizienten β_k, γ_i eines solchen Modells hat sich deshalb der Name "lineare Prädiktion" eingebürgert und die Koeffizienten werden als Prädiktorkoeffizienten bezeichnet (Begriff erstmals geprägt von WIENER, 1949). Wir wollen uns im folgenden mit dem bei der Untersuchung des Spracherzeugungsmodells wichtigen Spezialfall des allgemeinen Ansatzes (3.9) beschäftigen, bei dem alle

$$\gamma_i = 0 \quad , \quad 1 \leq i \leq q \qquad (3.10)$$

sind. Dieser Spezialfall wird auch als autoregressives Modell bezeichnet (DURBIN, 1960). Eine ausführliche Bibliographie zur linearen Prädiktion findet sich in MAKHOUL, 1975 und MARKEL & GRAY, 1976.

3.2.1 Grundprinzip der linearen Prädiktion

Das durch (3.9) beschriebene System kann mit der Einschränkung (3.10) wie in Bild 3.4 dargestellt werden. Das Ausgangssignal

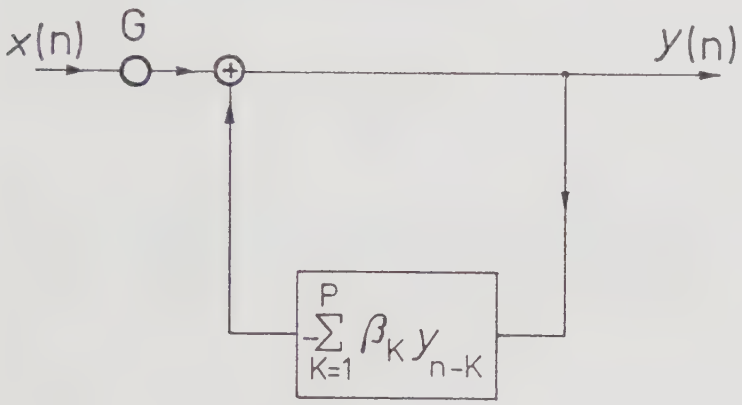


Bild 3.4 Rein rekursives Prädiktionsfilter

ergibt sich also als

$$y(n) = -\sum_{k=1}^{p} \beta_k \, y(n-k) + G \, x(n) \qquad (3.11)$$

Das Eingangssignal sei unbekannt. Daher läßt sich aus den p zurückliegenden Ausgangswerten nur das Signal

$$\tilde{y}(n) = -\sum_{k=1}^{p} \beta_k \, y(n-k) \qquad (3.12)$$

berechnen. Der Fehler zwischen dem tatsächlichen Wert y(n) und dem berechneten $\tilde{y}(n)$ ist dann

$$d(n) = y(n) - \tilde{y}(n) = y(n) + \sum_{k=1}^{p} \beta_k \, y(n-k) \qquad (3.13)$$

Die Koeffizienten β_k sollten natürlich Werte haben, die den entstehenden Fehler möglichst klein machen. Das Maß für den Fehler ist im Prinzip beliebig, aber um u.a. lineare Gleichungssysteme zu erhalten, wird häufig der mittlere quadratische Fehler minimiert. Aus (3.13) erhält man

$$D_q = \sum_n d^2(n) = \sum_n \left(y(n) + \sum_{k=1}^{p} \beta_k \, y(n-k) \right)^2 \qquad (3.14)$$

als den gesamten quadratischen Fehler, dabei werden die Grenzen der Summierung zunächst noch offen gelassen. Der Fehler D_q wird bezüglich der Koeffizienten β_i durch die Bedingung

$$\frac{\partial D_q}{\partial \beta_i} = 0 \qquad 1 \leq i \leq p \qquad (3.15)$$

minimiert. Mit (3.14) und (3.15) erhalten wir das Gleichungssystem

$$\sum_{k=1}^{p} \tilde{\beta}_k \sum_n y(n-k)\, y(n-i) = -\sum_n y(n)\, y(n-i) \qquad (3.16)$$

mit $1 \leq i \leq p$, mit Hilfe dessen wir die p unbekannten Koeffizienten $\tilde{\beta}_k$ berechnen können. Der minimale Fehler

$$D = \sum_n y^2(n) + \sum_{k=1}^{p} \tilde{\beta}_k \sum_n y(n)\, y(n-k) \qquad (3.17)$$

ergibt sich, wie gezeigt werden kann (MAKHOUL, 1975) durch Erweiterung von (3.14) und Einsetzen von (3.16). *

Bei der Festlegung der Summierungsgrenzen in (3.14, 3.16, 3.17) gibt es zwei interessante Fälle, die zu verschiedenen Methoden der Koeffizientenbestimmung führen. Summiert man über $-\infty < n < \infty$, so ergibt sich die Autokorrelationsmethode. Das Gleichungssystem (3.16) vereinfacht sich zu

* Da die Lösung von (3.16) eindeutig ist, benutzen wir im folgenden β statt $\tilde{\beta}$.

$$\underline{R}\ \underline{\beta} = -\underline{r} \tag{3.18}$$

wobei $\underline{\beta} = \{\beta_1, \beta_2, \ldots, \beta_p\}^T$, $\underline{r} = \{R(1), R(2), \ldots, R(p)\}^T$ und die $R_{ij} = R(i-j)$, $1 \leq i,j \leq p$, sind.

Die Funktion

$$R(i) = \sum_{n=-\infty}^{\infty} y(n)\ y(n-i) \tag{3.19}$$

ist die Autokorrelationsfunktion des Signals y(n). Da sie eine gerade Funktion, also

$$R(i) = R(-i)$$

ist, hat die sich ergebende Matrix $\underline{R}$ die Form einer symmetrischen Toeplitz-Matrix und wird als Autokorrelations-Matrix bezeichnet. Der minimale Fehler D (3.17) wird somit zu

$$D = R(o) + \underline{\beta}^T\ \underline{r} \tag{3.20}$$

In der Praxis ist das Signal y(n) meist nur in einem endlichen Intervall bekannt, oder man ist nur in einem endlichen Intervall an dem Signal interessiert. Durch Multiplikation mit einer Fensterfunktion W(n) erhält man das Signal

$$y'(n) = \begin{cases} y(n)\ W(n) & 0 \leq n \leq N-1 \\ 0 & \text{für alle anderen } n \end{cases} \tag{3.21}$$

das außerhalb des Intervalls $[0,N-1]$ Null ist. Die Autokorrelationsfunktion wird dann

$$R(i) = \sum_{n=0}^{N-i-1} y'(n)\ y'(n+i) \qquad |i| \leq N-1$$

$$R(i) = 0 \qquad |i| > N-1 \tag{3.22}$$

Auf die Frage der Form der Fensterfunktion werden wir später noch eingehen.

Will man den Fehler D_q nur in einem endlichen Intervall $0 \leq n \leq N$ minimieren, so ergibt sich die Kovarianzmethode. Das Gleichungssystem (3.16) vereinfacht sich zu

$$\underline{C}\,\underline{\beta} = \underline{c} \tag{3.23}$$

wobei der Vektor $\underline{c} = \{\phi(0,1), \phi(0,2), \ldots, \phi(0,p)\}^T$ und die Elemente der Matrix $C_{i,j} = \phi(i,j)$ $1 \leq i,j \leq p$ sind. Die Kovarianzfunktion $\phi(i,j)$ ist dabei durch

$$\phi(i,j) = \sum_{n=0}^{N-1} y(n-i)\, y(n-j) \qquad \begin{matrix} 0 \leq i \leq p \\ 0 \leq j \leq p \end{matrix} \tag{3.24}$$

definiert. Der minimale Fehler (3.17) wird dann zu

$$D = \phi(0,0) + \underline{\beta}^T \underline{c} \tag{3.25}$$

Da $\phi(j,i)$ gleich $\phi(i,j)$ ist, ist die Matrix $\underline{C}$ symmetrisch. Sie ist aber im allgemeinen keine Toeplitz-Matrix, da

$$\phi(i+1,j+1) = \phi(i,j)+y(-i-1)y(-j-1)-y(N-1-i)y(N-1-j) \tag{3.26}$$

ist (MAKHOUL, 1975). An (3.26) erkennt man, daß zur Berechnung aller Elemente der Matrix das Signal $y(n)$ für $-p \leq n \leq N-1$ bekannt sein muß.

Sind nun die Signalwerte $y(n)$ Abtastwerte eines stochastischen Prozesses, so sind es auch die Fehler $d(n)$ (3.13). Somit muß man den Erwartungswert des quadratischen Fehlers minimieren. Das Gleichungssystem (3.16) wird also zu

$$\sum_{k=1}^{p} \beta_k\, E\{y(n-k)\, y(n-i)\} = -\, E\{y(n)\, y(n-i)\} \tag{3.27}$$

mit $1 \leq i \leq p$. Der minimale Fehler D ergibt sich dann zu

$$D = E\{y^2(n)\} + \sum_{k=1}^{p} \beta_k \, E\{y(n) \; y(n-k)\} \qquad (3.28)$$

Bei stationären, ergodischen Prozessen ist

$$E\{y(n-k) \; y(n-i)\} = R(i-k) \; , \qquad (3.29)$$

also die Autokorrelationsfunktion. Bei der betrachteten Klasse von Prozessen läßt sich die Autokorrelationsfunktion als zeitlicher Mittelwert berechnen (BENDAT & PIERSOL, 1971), so daß wir die durch (3.22) definierten Werte benutzen können. Diese Werte werden üblicherweise durch die Anzahl N der Signalwerte dividiert. Da dies aber die Lösung des Gleichungssystems nicht beeinflußt, läßt man diese Operation häufig weg. Für einen stationären, ergodischen Prozeß ergibt sich also der gleiche Lösungsgang wie bei der Autokorrelationsmethode. Für den Fall eines nichtstationären Prozesses ist

$$E\{y(n-k) \; y(n-i)\} = R(n-k,n-i) \; , \qquad (3.30)$$

d.h. die Autokorrelationsfunktion ist auch noch eine Funktion des Abtastzeitpunktes n. MAKHOUL, 1975 zeigt nun, daß für eine bestimmte Klasse der nichtstationären Prozesse, die als lokal stationär bezeichnet werden und zu denen das Sprachsignal und EEG Signale gezählt werden können, die Berechnung der Koeffizienten β_k identisch mit der des Kovarianzverfahrens ist.

Da das Eingangssignal x(n) unbekannt ist, ist es natürlich auch nicht sinnvoll, den Verstärkungsfaktor G (3.11) zu bestimmen. Schreibt man allerdings (3.13) als

$$y(n) = - \sum_{k=1}^{p} \beta_k \; y(n-k) + d(n) \; , \qquad (3.31)$$

so erkennt man, daß das Eingangssignal x(n) proportional zum Fehlersignal d(n) sein muß, wenn es das Ausgangssignal y(n) erzeugen soll. Bei einem beliebigen Eingangssignal läßt sich

mit dem Verstärkungsfaktor G die Energie des Eingangssignals an die des Fehlersignals anpassen. Die Energie des Signals G x(n) ist dann gleich dem minimalen quadratischen Fehler D.

Bei dem in Bild 3.2 dargestellten Modell der Spracherzeugung werden zwei Eingangssignale betrachtet. Das erste ist

$$x(n) = \delta(n) \tag{3.32}$$

wobei

$$\delta(n) = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases} \tag{3.33}$$

der Einheitsimpuls ist. Das Ausgangssignal des in Bild 3.4 dargestellten Systems auf einen mit dem Verstärkungsfaktor G gewichteten Einheitsimpuls ist die Impulsantwort

$$h(n) = 0 \quad ; \quad n < 0$$

$$h(n) = -\sum_{k=1}^{p} \beta_k \, h(n-k) + G \, \delta(n) \quad ; \quad n \geq 0 \tag{3.34}$$

Es wurde gezeigt (MAKHOUL, 1973), daß für die Autokorrelationsfunktion dieser Impulsantwort gilt

$$\hat{R}(i) = -\sum_{k=1}^{p} \beta_k \, \hat{R}(i-k) \qquad 1 \leq i \leq \infty \tag{3.35}$$

$$\hat{R}(o) = -\sum_{k=1}^{p} \beta_k \, \hat{R}(k) + G^2 \tag{3.36}$$

Mit der Bedingung, daß die gesamte Energie der Impulsantwort h(n) gleich der des Ausgangssignals y(n) sein soll, ergibt sich für die Autokorrelationsfunktion R(i) des Signals y(n)

$$R(o) = \hat{R}(o) \tag{3.37}$$

Da sich (3.35) und (3.18) nur im Wertebereich des Index i unterscheiden, schließt MAKHOUL, daß

$$R(i) = \hat{R}(i) \qquad 0 \leq i \leq p \qquad (3.38)$$

ist, d.h. die ersten p + 1 Werte der Autokorrelationsfunktion der Impulsantwort h(n) des Systems sind gleich den entsprechenden Werten der Autokorrelationsfunktion des Ausgangssignals y(n).

Das zweite in Frage kommende Eingangssignal x(n) ist eine Folge von unkorrelierten Abtastwerten (weißes Rauschen) mit dem Mittelwert Null. Für die Autokorrelationsfunktion des zugehörigen Ausgangssignals eines Systems entsprechend Bild 3.4 ist gezeigt worden (BOX & JENKINS, 1970), daß die Beziehungen (3.35) und (3.36) ebenfalls gelten. Da auch in diesem Fall die Energie des Ausgangssignals gleich der Energie des Signals y(n) sein soll, gelten auch (3.37) und (3.38).

Zusammenfassend kann man sagen, daß die Autokorrelationsfunktion des Ausgangssignales und die Koeffizienten β_k des in Bild 3.4 dargestellten Systems sowohl bei der Eingangsgröße "Einheitsimpuls δ", als auch bei der Eingangsgröße "weißes Rauschen" über das Gleichungssystem (3.18) miteinander verknüpft sind. Da die beiden Eingangssignale die gleiche Autokorrelationsfunktion haben, ist dieses Ergebnis auch plausibel. Aufgrund unseres Spracherzeugungsmodells in Bild 3.2 können wir nun mit demselben Algorithmus sowohl stimmhafte als auch stimmlose Sprache analysieren und die entsprechenden Koeffizienten β_k berechnen.

3.2.2 Lösungsmethoden

Bei beiden angegebenen Lösungswegen (3.18) und (3.23) muß ein Gleichungssystem mit p Gleichungen und p Unbekannten gelöst werden. Neben den Standardmethoden für diese Aufgabe ergeben sich aufgrund der Eigenschaften der Gleichungssysteme Lösungsverfahren, die weniger Rechenoperationen und Speicherplatz benötigen. Nutzt man bei der Lösung des Gleichungssystems (3.18) des Autokorrelationsverfahrens die Toeplitzform der Matrix R

aus, so bietet sich ein von DURBIN, 1960 vorgestellter rekursiver Algorithmus an, der diese Aufgabe sehr effektiv löst. Die Rekursionsformel sieht dabei folgendermaßen aus

$$D_o = R(o) \tag{3.39a}$$

$$k_1 = -R(1)/D_o \tag{3.39b}$$

$$k_i = -\left[R(i) + \sum_{j=1}^{i-1} \beta_j^{(i-1)} R(i-j)\right] D_{i-1}^{-1}, \quad 2 \le i \le p \tag{3.39c}$$

$$\beta_i^{(i)} = k_i \tag{3.39d}$$

$$\beta_j^{(i)} = \beta_j^{(i-1)} + k_i \beta_{i-j}^{(i-1)}, \quad 1 \le j \le i-1 < p \tag{3.39e}$$

$$D_i = (1 - k_i^2) D_{i-1}, \tag{3.39f}$$

und wird für i = 1 bis p gelöst. Die endgültige Lösung ist dann

$$\beta_j = \beta_j^{(p)} \qquad 1 \le j \le p. \tag{3.39g}$$

Zu beachten ist bei diesem Lösungsalgorithmus, daß auch alle Prädiktorkoeffizienten β_k für alle Systeme niedrigerer Ordnung als p berechnet werden. Außerdem wird dabei der jeweilige minimale Gesamtfehler D_i berechnet. Es ist gezeigt worden (LEVINSON, 1946), daß

$$0 \le D_i \le D_{i-1} \tag{3.40}$$

gilt. Dies ermöglicht es, die Rekursion bei Unterschreiten einer gegebenen Fehlerschranke abzubrechen und somit jeweils ein System mit der jeweils minimal nötigen Ordnung p zu erhalten.

Die Lösung des Gleichungssystems (3.23) der Kovarianzmethode

$$\underline{C}\ \underline{\beta} = \underline{c}$$

läßt sich aufgrund der Symmetrie, und da die Matrix positiv definit (oder semidefinit) ist, mit Hilfe der Cholesky Zerlegung (ATAL & HANAUER, 1971) durchführen. Wir wollen auf diese Lösungsmethode nicht weiter eingehen, da wir die Kovarianzmethode aus Gründen, die wir noch darstellen werden, nicht benutzt haben.

Während bei der Korrelationsmethode und bei der Kovarianzmethode die Lösung in zwei Schritten, Berechnung von Matrixelementen und Auflösung eines linearen Gleichungssystems, geschieht, werden diese beiden Schritte bei einer dritten Klasse von Methoden miteinander zu rekursiven Algorithmen verbunden (MAKHOUL, 1977). Sie werden in der englischen Literatur als "Lattice" Methoden bezeichnet. Bildet man aus der Differenzengleichung (3.11) die Übertragungsfunktion

$$H(z) = \frac{Y(z)}{X(z)} = \frac{G}{1 + \sum_{k=1}^{p} \beta_k \, z^{-k}} \qquad (3.41)$$

so wird das Nennerpolynom

$$N(z) = 1 + \sum_{k=1}^{p} \beta_k \, z^{-k} \qquad (3.42)$$

als Übertragungsfunktion des inversen Filters bezeichnet. Es stellt, wie man an (3.13) sieht, die Übertragungsfunktion zwischen Y(z) und dem Fehler D(z) dar und läßt sich durch die in Bild 3.5 dargestellte Struktur realisieren (ITAKURA & SAITO, 1971). Die Prädiktorkoeffizienten β_k hängen über die Rekursionsformeln (3.39c,d) mit den Koeffizienten k_i zusammen. Die Koeffizienten k_i können bei dieser Klasse von Methoden durch Minimierung einer Norm des Vorwärtsprädiktionsfehlers $f_m(n)$ oder des Rückwärtsprädiktionsfehlers $b_m(n)$ berechnet werden. Durch verschiedene Verknüpfung der Fehler ergeben sich mehrere unterschiedliche Verfahren zur Berechnung der Koeffizienten k_i (MAKHOUL, 1977).

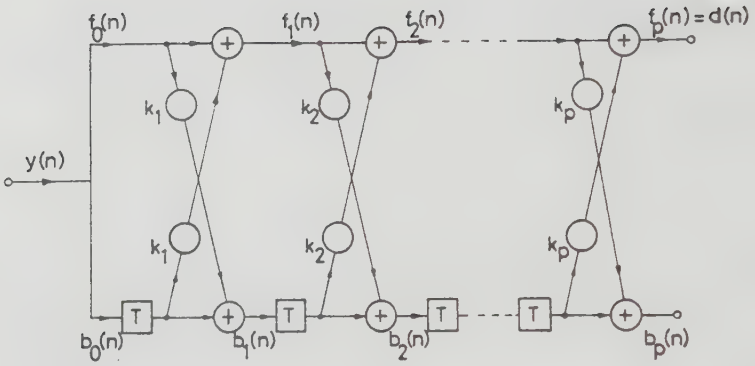


Bild 3.5 Digitale Implementierung der "Lattice" Methode

3.2.3 Zusammenhang zwischen Prädiktor- und Adaptorkoeffizienten

Formen wir die Übertragungsfunktion H(z) (3.41) zu

$$H(z) = \frac{Y(z)}{X(z)} = \frac{z^p \quad G}{\sum_{k=o}^{p} \beta_{p-k} \; z^k} \tag{3.43}$$

mit $\beta_o = 1$ um, so erkennen wir, daß sie die gleiche Form wie die Übertragungsfunktion (2.36) des Leitungsmodells hat. Wir benötigen also ein Verfahren, um aus den Prädiktorkoeffizienten β_k die Adaptorkoeffizienten α_i zu berechnen. Durch die von uns gewählte Beschreibung mit Transfermatrizen bzw. kanonischen Polynomen können wir diesen Zusammenhang, im Gegensatz zu dem bisher in der Literatur beschriebenen (z.B. ISOMICHI, 1974), recht einfach herstellen. Dazu gehen wir von dem einfachsten Fall eines Prädiktors erster Ordnung aus. Die Übertragungsfunktion ist

$$P^{(1)}(z) = \frac{z \cdot G}{z + \beta_1^{(1)}} \tag{3.44}$$

die entsprechende Übertragungsfunktion des Leitungsmodells bestehend aus zwei Leitungsstücken und einem Adaptor ist nach (2.36)

$$L^{(1)}(z) = \frac{z^2 (1 - \alpha_1)}{z^2 + \alpha_S \alpha_1 z + \alpha_G \alpha_1 z + \alpha_G \alpha_S} \tag{3.45}$$

Um nun in beiden Fällen abgesehen vom Verstärkungsfaktor die gleiche Übertragungsfunktion zu erhalten, gibt es, wie wir sehen, unbeschränkt viele Möglichkeiten, da wir neben der zwingenden Bedingung, daß einer der Koeffizienten α_G und α_S Null sein muß, den anderen frei wählen können. Wir wählen willkürlich

$$\alpha_G = 1$$

und (3.46)

$$\alpha_S = 0$$

dann ergibt sich

$$\beta_1^{(1)} = \alpha_1$$

Allgemein können wir die Übertragungsfunktion L(z) entsprechend (2.36) mit der Bedingung (3.46) als

$$L^{(i)}(z) = \frac{f^{(i)}(z)}{g^{(i)}(z) - h^{(i)}(z)} \tag{3.47}$$

schreiben, wobei der hochgestellte Index i wie schon bei (3.44) und (3.45) die Ordnung des Prädiktorfilters bzw. hier die Anzahl der Adaptoren angibt. Einer Erhöhung der Ordnung entspricht beim Leitungsmodell dem Hinzufügen eines Adaptors und eines Leitungselementes, also der Multiplikation der Transfer-

matrix $\underline{T}^{(i)}$ des Systems mit i Adaptoren mit

$$\underline{T}_{i+1} = \frac{1}{z(1 - \alpha_{i+1})} \begin{pmatrix} 1 & -\alpha_{i+1} z \\ -\alpha_{i+1} & z \end{pmatrix} \tag{3.48}$$

um mit

$$\underline{T}^{(i+1)} = \underline{T}^{(i)} \underline{T}_{i+1} \tag{3.49}$$

die Transfermatrix des Systems mit i + 1 Adaptoren zu erhalten. War das Nennerpolynom von $L^{(i)}(z)$

$$N^{(i)}(z) = g^{(i)}(z) - h^{(i)}(z) \ , \tag{3.50}$$

so ist das Nennerpolynom von $L^{(i+1)}(z)$ wegen (3.47) bis (3.49)

$$N^{(i+1)}(z) = -\alpha_{i+1} \, z h_*^i + z g^i + \alpha_{i+1} \, z g_*^i - z h^i \tag{3.51}$$

oder ausgedrückt durch die Polynome $N^{(i)}(z)$ und $N_*^{(i)}(z)$

$$N^{(i+1)}(z) = z N^{(i)}(z) + \alpha_{i+1} z \, N_*^{(i)}(z) \tag{3.52}$$

Setzen wir also das Polynom als

$$N^{(i-1)} = \sum_{\nu=1}^{i} z^\nu \, \beta_{i-\nu}^{(i-1)} \tag{3.53}$$

an, so ergibt sich mit (3.52) das Polynom

$$N^{(i)}(z) = z \sum_{\nu=1}^{i} z^\nu \, \beta_{i-\nu}^{(i-1)} + \alpha_i \, z \, z^i \sum_{\nu=1}^{i} \frac{1}{z^\nu} \, \beta_{i-\nu}^{(i-1)}$$

$$= \sum_{\nu=1}^{i} z^{\nu+1} \beta_{i-\nu}^{(i-1)} + \alpha_i \sum_{\nu=1}^{i} z^{i+1-\nu} \beta_{i-\nu}^{(i-1)} \qquad (3.53)$$

Setzen wir ferner

$$\mu = \nu + 1,$$

so ergibt sich

$$N^{(i)}(z) = \sum_{\mu=2}^{i+1} z^{\mu} \beta_{i+1-\mu}^{(i-1)} + \alpha_i \sum_{\mu=1}^{i} z^{\mu} \beta_{\mu-1}^{(i-1)}$$

$$= \alpha_i z + \beta_o^{(i-1)} z^{i+1} + \sum_{\mu=2}^{i} \left(\beta_{i+1-\mu}^{(i-1)} + \alpha_i \beta_{\mu-1}^{(i-1)} \right) z^{\mu} \qquad (3.54)$$

Ein Koeffizientenvergleich mit

$$N^{(i)}(z) = \sum_{\nu=1}^{i+1} z^{\nu} \beta_{i+1-\nu}^{(i)}$$

ergibt

$$\beta_i^{(i)} = \alpha_i \qquad (3.55a)$$

$$\beta_o^{(i)} = \beta_o^{(i-1)} \qquad (3.55b)$$

$$\beta_{i+1-\nu}^{(i)} = \beta_{i+1-\nu}^{(i-1)} + \alpha_i \beta_{\nu-1}^{(i-1)} \quad 2 \le \nu \le i \qquad (3.55c)$$

Setzt man in (3.55c) noch $i+1-\nu = j$, so ergibt sich mit

$$\beta_j^{(i)} = \beta_j^{(i-1)} + \alpha_i \beta_{i-j}^{(i-1)} \quad 1 \le j \le i-1 \qquad (3.56)$$

die gleiche Rekursionsformel wie bei der Auflösung des Gleichungssystems mit Hilfe der DURBIN Rekursion (3.39d). Da, wie

man an (3.44) sieht,

$$\beta_o^{(1)} = 1$$

ist, sind nach (3.55b) alle $\beta_o^{(i)} = 1$ und damit haben wir gezeigt, daß das Leitungsmodell mit der Bedingung (3.46) und den Adaptorkoeffizienten

$$\alpha_i = k_i$$

bis auf den Verstärkungsfaktor die gleiche Übertragungsfunktion hat wie das Prädiktorfilter mit den Koeffizienten $\beta_k^{(p)}$ (k=0,1,..,p). Mit (3.55a) und (3.56) haben wir auch eine Rekursionsformel, um aus bekannten α_i die entsprechenden β_i zu berechnen, indem wir i die Werte 1,2,...,p annehmen lassen. Sind die Prädiktorkoeffizienten $\beta_i^{(p)}$ bekannt, so lassen sich aus (3.55a) und (3.56) die Rekursionsformeln

$$\alpha_i = \beta_i^{(i)} \qquad (3.57a)$$

$$\beta_j^{(i-1)} = \frac{\beta_j^{(i)} - \alpha_i \beta_{i-j}^{(i)}}{1 - \alpha_i^2} \qquad 1 \le j \le i-1 \qquad (3.57b)$$

ermitteln, wobei i die Werte p,p-1,...,1 annimmt, mit der wir die unbekannten Adaptorkoeffizienten berechnen können. Um einzusehen, daß (3.57b) aus(3.56) folgt, ersetzen wir zunächst in (3.56) den Index j durch i-j. Mit Hilfe des sich hierbei ergebenden Zusammenhangs kann dann die Größe $\beta_{i-j}^{(i-1)}$ in (3.56) eliminiert werden. Eine Auflösung der resultierenden Gleichung nach $\beta_j^{(i-1)}$ ergibt schließlich (3.57b).

Auf Grund unserer Ergebnisse aus Kapitel 2 wissen wir, daß bei Vertauschung der Werte für α_G und α_S bei gleichzeitiger Änderung der Reihenfolge der Adaptorkoeffizienten α_i sich die Übertragungsfunktion des Leitungsmodelles nicht ändert. Diese beiden willkürlichen Annahmen über die Randbedingungen, die allerdings zu dem beschriebenen einfachen Zusammenhang zwischen den Adaptorkoeffizienten und den Prädiktorkoeffizienten führen, sind die in Kap. 3.1.2 erwähnten Randbedingungen von ATAL, 1970

bzw. WAKITA, 1973. Sie gewinnen mit (3.1) aus den Adaptorkoeffizienten einen Satz Reflektionskoeffizienten r_i, den sie der tatsächlichen Querschnittsflächenfunktion zuordnen.

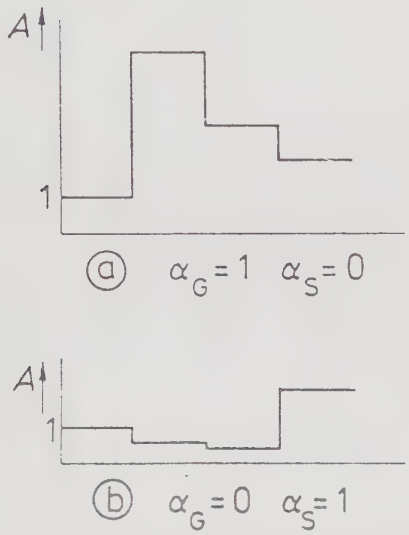


Bild 3.6 Querschnittsflächenfunktionen von akustischen Rohren mit identischer Übertragungsfunktion

Bild 3.6 zeigt als Beispiel die beiden, aus dem Prädiktorkoeffizientensatz $\beta_1 = 0{,}55$; $\beta_2 = -0{,}35$; $\beta_3 = -0{,}2$ mit den Randbedingungen von ATAL, Teil a, und den Randbedingungen von WAKITA, Teil b, gewonnenen Querschnittsfunktionen. Da die zugehörigen Reflektionskoeffizientensätze $r_1 = -0{,}\bar{6}$; $r_2 = 0{,}25$; $r_3 = 0{,}2$ bzw. $r_1 = 0{,}2$; $r_2 = 0{,}25$; $r_3 = -0{,}\bar{6}$ nur jeweils die Verhältnisse der benachbarten Querschnittsflächen beschreiben, wurde die erste Querschnittsfläche gleich Eins gewählt. Wie man bei (3.45) sieht, ist es natürlich auch möglich, α_G bzw. α_S einen Wert aus dem Intervall $[-1, 1]$ zuzuordnen. Dann gilt natürlich der einfache rekursive Zusammenhang zwischen Prädiktor- und Adaptorkoeffizienten nicht mehr, vielmehr müssen wir durch Koeffizientenvergleich den jeweils unbekannten Koeffizientensatz berechnen. Dabei sind dann beliebig viele unter-

schiedliche Koeffizientensätze möglich, die aber alle die gleiche Übertragungsfunktion haben. Der als Beispiel angeführte Prädiktorkoeffizientensatz ergibt zum Beispiel mit $\alpha_G = 0{,}75$, $\alpha_S = 0$ die Reflektionskoeffizienten $r_1 = -0{,}897$, $r_2 = 0{,}194$, $r_3 = 0{,}2\overline{6}$. Das entspricht, wieder mit der Normierung der ersten Querschnittsfläche auf 1, der in Bild 3.7

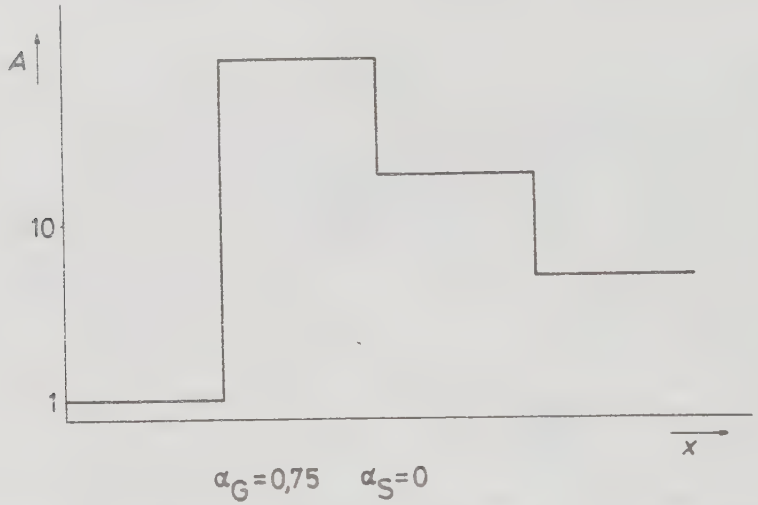


Bild 3.7 Querschnittsflächenfunktion eines akustischen Rohres mit gleicher Übertragungsfunktion wie Rohre nach Bild 3.6

dargestellten Querschnittfunktion. Alle drei in Bild 3.6 und 3.7 dargestellten Querschnittsverläufe haben mit den entsprechenden Randbedingungen die gleiche Übertragungsfunktion. Um den tatsächlichen Querschnittsverlauf zu erhalten, muß man die jeweils vorhandenen Randbedingungen kennen. Da dies nur relativ ungenau der Fall ist, scheinen uns, wie wir schon ausgeführt hatten, die Verfahren von ATAL bzw. WAKITA nicht sehr geeignet, den tatsächlichen Querschnittsverlauf zu ermitteln.

Für das in Bild 3.2 dargestellte Modell der Spracherzeugung reicht die Kenntnis der Übertragungsfunktion S(z) aus, so daß wir bei Kenntnis der Prädiktorkoeffizienten mit den Rekursionsformeln oder direkt mit dem Durbin Algorithmus die Adaptorkoeffizienten aus dem Sprachsignal gewinnen können. Denn für die

Realisierung der Übertragungsfunktion ist es natürlich sinnvoll, einem der beiden Randbedingungsfälle $\alpha_G = 1$, $\alpha_S = 0$ oder $\alpha_G = 0$, $\alpha_S = 1$ zu benutzen.

3.2.4 Gewähltes Verfahren und praktische Gewinnung der Koeffizienten

Bei der Auswahl des Verfahrens spielt sowohl die Frage des Rechenaufwandes, als auch die Frage der Stabilität des Filters eine Rolle. Beim Rechenaufwand ergibt sich, daß die Autokorrelationsmethode den geringsten Speicherplatz und die kürzeste Rechenzeit benötigt (MAKHOUL, 1977). Das durch die Übertragungsfunktion H(z) (3.43) beschriebene System ist stabil, wenn die Nullstellen des Nennerpolynoms N(z) im Einheitskreis liegen. Die Kenntnis der Polynomkoeffizienten β_k erlaubt es nur durch aufwendige Verfahren der Nullstellenbestimmung, die Stabilität festzustellen. Für die Adaptorkoeffizienten des Leitungsmodells, das die Übertragungsfunktion H(z) realisiert, gilt allerdings aufgrund von (2.12) die einfache Bedingung

$$|\alpha_i| \leq 1 \qquad i = 1,2,\ldots,p \tag{3.58}$$

zur Sicherstellung der Stabilität. Mit Hilfe der rekursiven Beziehung zwischen den beiden Koeffizientensätzen ergibt sich also ein relativ einfacher Stabilitätstest für bekannte Prädiktorkoeffizienten β_k. Da bei der Kovarianzmethode die Prädiktorkoeffizienten direkt berechnet werden, muß man, da eine Stabilitätsgarantie aufgrund der Berechnungsmethode nicht gegeben werden kann, jeden Koeffizientensatz auf Stabilität überprüfen. LEVINSON, 1946 hat, wie schon erwähnt, gezeigt, daß der jeweilige Gesamtfehler D_i bei der Autokorrelationsmethode die Bedingung

$$0 \leq D_i \leq D_{i-1} \tag{3.59}$$

erfüllt. Aufgrund der Beziehung

$$D_i = (1 - \kappa_i^2)\, D_{i-1}$$

im Rekursionsalgorithmus von DURBIN folgt sofort, daß

$$|k_i| \leq 1$$

ist. Da diese Koeffizienten k_i unter den bereits genannten Randbedingungen den Adaptorkoeffizienten α_i entsprechen, garantiert die Gewinnung der Koeffizienten nach der Autokorrelationsmethode ein stabiles Filter. Auch die meisten "Lattice" Methoden garantieren ein stabiles Filter (MAKHOUL, 1977), wobei die Stabilität auch bei Berechnung mit der endlichen Wortlänge eines Digitalrechners gewährleistet ist. In der Praxis konnten wir allerdings auch beim Autokorrelationsverfahren bei der Berechnung auf dem Laborrechner HP 21-MX in Gleitkommaarithmetik - 23 bit Mantissen - und 7 bit Exponentenwortlänge - bis jetzt keine instabilen Lösungen feststellen. Aufgrund der Speicherplatzersparnis und der geringeren Rechenzeit haben wir uns für das Autokorrelationsverfahren mit dem rekursiven Lösungsalgorithmus nach DURBIN entschieden.

Das Vokaltraktmodell ist, da es nur für den Ausbreitungsbereich ebener Wellen Gültigkeit hat, auf den Frequenzbereich bis 4 kHz beschränkt. Diese Beschränkung fällt bei dem in Bild 3.2 dargestellten Modell der Spracherzeugung weg, da wir das Sprachsignal approximieren und damit nicht unbedingt die physikalischen und physiologischen Erzeugungsmechanismen natürlicher Sprache berücksichtigen. Da eine Berücksichtigung von Frequenzen oberhalb von 5 kHz keinen wesentlichen Einfluß mehr auf die Qualität der Sprache hat, beschränken wir uns auf den Frequenzbereich bis 5 kHz. Zur Berechnung der Adaptorkoeffizienten werden die Sprachproben in digitaler Form auf der Magnetplatte bzw. dem Magnetband des Laborrechners gespeichert. Die zur Begrenzung des analogen Signals auf Frequenzen unterhalb der halben Abtastfrequenz benutzte Filterkennlinie hat eine Butterworth Charakteristik.Da es aufgrund des maximal flachen Verlaufes des Durchlaßbereiches einen relativ breiten Übergangsbereich bis zur erforderlichen Sperrdämpfung von 60 dB hat und nur bis zur halben Grenzfrequenz die Forderung nach linearer Phase genügend genau erfüllt, werden die Sprachproben mit dem in Bild 3.8 dargestellten System aufgezeichnet. Das Mikro-

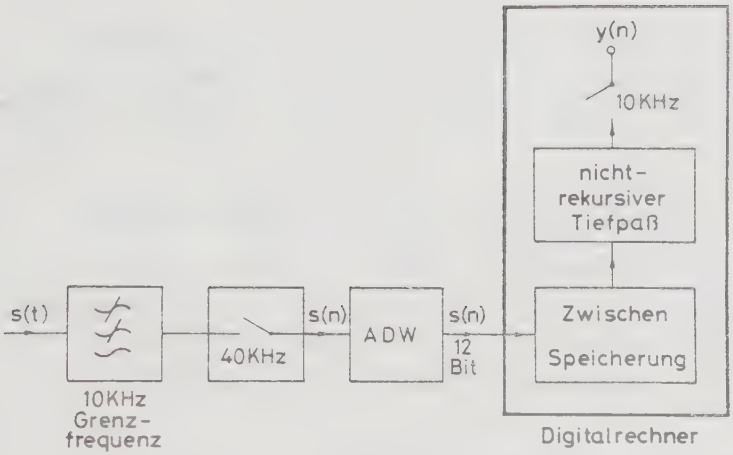


Bild 3.8 Blockschaltbild des Aufbaues zur Aufzeichnung der Sprachproben

phon steht, um Einflüsse des Raumes auf das Sprachsignal zu vermeiden, im reflektionsarmen Meßraum. Die Abtastwerte des mit 40 kHz abgetasteten Signals werden nach der Analog-Digital Wandlung in 12 bit lange Datenwerte im 16 bit Integerformat des Rechners auf der Magnetplatte zwischengespeichert. Die Bandbegrenzung durch ein Tiefpaßfilter mit linearem Phasengang auf 5 kHz und die Herabsetzung der Abtastfrequenz auf 10 kHz wird in zwei Stufen durch ein programmiertes nichtrekursives Tiefpaßfilter (Studienarbeit HENNEKE, 1977) durchgeführt. Der Dämpfungsverlauf wird durch eine modifizierte Fourier-Approximation angenähert (SCHÜSSLER, 1973). Nach der ersten Filterung mit einem Filter 50. Ordnung, der Grenzfrequenz 10 kHz und einer Reduzierung der Abtastfrequenz auf 20 kHz ergibt die zweite Filterung mit einem Filter 200. Ordnung einen Dämpfungsverlauf, der unterhalb 4,64 kHz kleiner als 0,01 dB ist und bei 5 kHz 60 dB Sperrdämpfung erreicht. Um diese Filterung zu erzielen, müssen die Berechnungen natürlich alle in Gleitkommaarithmetik erfolgen, die gespeicherten Abtastwerte werden deshalb zunächst entsprechend gewandelt. Aus dem so vorbereiteten Sprachsignal können dann mit dem Programm LPCAN (Flußdiagramm im Anhang) die Adaptorkoeffizienten berechnet werden. Dabei

sind sowohl die Länge des gleitenden Analysefensters, der Vorschub des Fensters, als auch die Ordnung der Analyse frei wählbar. Außerdem können zwei verschiedene Fensterfunktionen benutzt werden.

3.2.5 Parametereinflüsse auf die lineare Prädiktion

Im folgenden wollen wir die Einflüsse der verschiedenen Parameter auf die lineare Prädiktion untersuchen, wobei wir uns auf das von uns gewählte Autokorrelationsverfahren beschränken wollen. Die gewonnenen experimentellen Aussagen wurden für natürliche Sprache gewonnen.

3.2.5.1 Die Ordnung der Prädiktion

Ein Maß für die Güte der Prädiktion in Abhängigkeit von der Prädiktorordnung p ist sicherlich die Größe des minimalen quadratischen Fehlers D (3.20), wobei man zweckmäßigerweise auf die Energie des Signals normiert und dann

$$V(p) = \frac{D(p)}{R(o)} \tag{3.60}$$

als normierten minimalen quadratischen Fehler erhält.

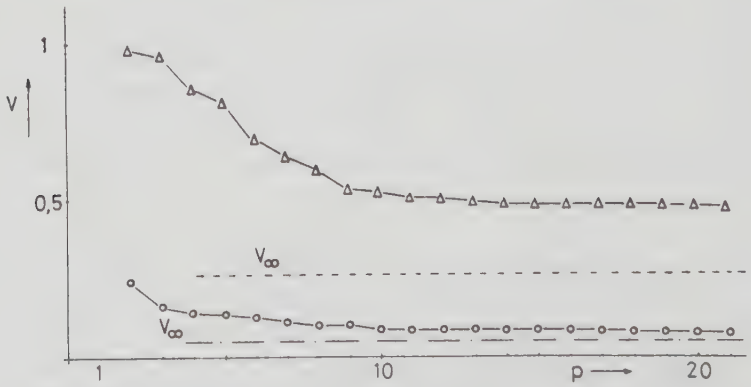


Bild 3.9 Normierter minimaler quadratischer Fehler V als Funktion der Prädiktionsordnung p.
Δ ≙ stimmlose Sprache; o ≙ stimmhafte Sprache

Bild 3.9 zeigt diese Funktion für die beiden in Bild 3.10

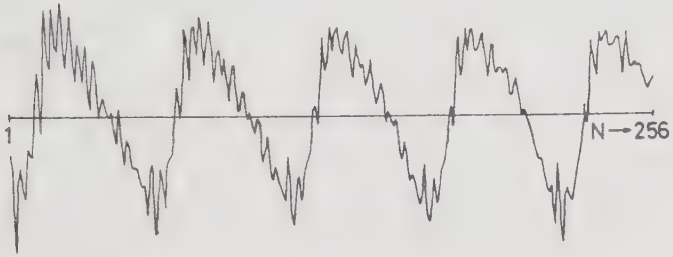


Bild 3.10 Für Bild 3.9 benutzte Proben stimmhafter und stimmloser Sprache

dargestellten Sprachproben. Nach (3.59) muß V(p) für größer werdende p kleiner werden oder konstant bleiben. Für $p \to \infty$ erreicht V(p) also den minimalen Wert V_∞. Da dieser Wert, wie man auch schon an Bild 3.9 erkennen kann, nicht notwendigerweise Null ist, sagt der absolute Wert von V(p) allein auch noch nichts über die Güte der Prädiktion aus, er muß immer im Vergleich mit dem Grenzwert V_∞ gesehen werden. Die Bestimmung dieses Grenzwertes ergibt sich bei der Betrachtung der linearen Prädiktion im Frequenzbereich. Aus (3.14) ergibt sich für den Fehler eines Analyseintervalls mit N Abtastwerten

$$D = \sum_{n=0}^{N-1} d^2(n) = \sum_{n=0}^{N-1} \left[\sum_{k=0}^{p} \beta_k \, y(n-k) \right]^2$$

Die Anwendung des Parsevalschen Theorems für den diskreten Fall (TRETTER, 1976) ergibt

$$D = \frac{1}{N} \sum_{l=0}^{N-1} |\tilde{D}(l)|^2$$

bzw.

$$D = \frac{1}{N} \sum_{l=0}^{N-1} |\tilde{Y}(l)|^2 \cdot |\tilde{N}(l)|^2 \qquad (3.61)$$

als Fehler. Dabei sind $\tilde{D}(l)$, $\tilde{Y}(l)$ und $\tilde{N}(l)$ die diskreten Fourier Transformierten der Folgen d(n), y(n) bzw. β_k. Da die Elemente der Folge β_k für $k \notin \{0,1,2,..,p\}$ gleich Null sind und $p < N$ ist, gilt

$$\tilde{N}(l) = N\left(e^{j\frac{2\pi}{N}l}\right)$$

wobei N(z) die in 3.42 angegebene Übertragungsfunktion des inversen Filters ist. Somit können wir auch

$$D = \frac{G^2}{N} \sum_{l=0}^{N-1} \frac{|\tilde{Y}(l)|^2}{\left|H\left(e^{j\frac{2\pi}{N}l}\right)\right|^2} \qquad (3.62)$$

schreiben, wobei H(·) die in (3.41) angegebene Funktion darstellt. Der Fehler D beschreibt also die Güte der Approximation des Betragsspektrums des Signals durch das Betragsspektrum des Modells. Da wegen (3.38) die ersten p + 1 Autokorrelationskoeffizienten des Signals und der Impulsantwort gleich sind, gilt für $p \to \infty$, da die Betragsspektren die Fouriertransformation der Autokorrelationsfunktion sind,

$$|Y(l)|^2 = |H(l)|^2 \text{ für } p \to \infty \ , \quad N \to \infty \ . \qquad (3.63)$$

Die Identität der Betragsspektren wirft natürlich gleich die Frage nach der Beziehung zwischen den Spektren auf. Da aufgrund der Struktur des Modells H(z) immer minimalphasig ist, gilt die Identität nur dann, wenn auch Y(z) minimalphasig ist. Dabei gilt, daß, wenn die Ordnung des das Signal y(n) erzeugenden Systems p_o ist, die Identität für alle $p \geq p_o$ besteht. Da das Signal y(n), das wir analysieren wollen, natürliche Sprache ist, die im allgemeinen kein Signal eines minimalphasigen Systems ist, gilt im allgemeinen nur (3.63). Wie man nun durch Vergleich von (3.20) und (3.36) sieht, gilt bei der Anregung durch einen Einheitsimpuls

$$D = G^2 \qquad (3.64)$$

um die gleiche Energie im Signal y(n) und in der Impulsantwort h(n) zu haben. Man kann den minimalen quadratischen Fehler aber auch mit Hilfe des nullten Cepstrum-Koeffizienten definieren (MAKHOUL, 1973). Allgemein ergibt sich der n-te Cepstrum-Koeffizient zu

$$c(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[\ln \; X(e^{j\omega})\right] e^{j\omega n} \, d\omega \qquad (3.65)$$

wobei X die zweiseitige z-Transformation der Folge x(n) ist,d.h.

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x(n) \; e^{-j\omega n} \qquad (3.66)$$

Für n = 0 läßt sich (3.65) zu

$$c(o) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln \, |X(e^{j\omega})| \, d\omega \qquad (3.67)$$

vereinfachen. Erfüllt die Folge x(n) die Bedingung

$$x(n) = 0 \quad \text{für} \quad n < 0$$

und ist x(n) minimalphasig, so läßt sich zeigen (GOLD & RADER, 1969), daß

$$c(o) = \ln \, (x(o)) \qquad (3.68)$$

ist. Für unser Spracherzeugungsmodell mit der Übertragungsfunktion

$$H(z) = \frac{G}{1 + \sum_{k=1}^{p} \beta_k \; z^{-k}}$$

ergibt sich also der durch den Index M gekennzeichnete Cepstrum-Koeffizient

$$c_M(o) = \ln G \tag{3.69}$$

und mit (3.69) und (3.64)

$$D = e^{2c_M(o)} \tag{3.70}$$

Zur Bestimmung von V_∞ brauchen wir also wegen (3.63) nur den Cepstrum-Koeffizienten c(o) des Eingangssignals y(n) zu bestimmen. Da wir immer nur Folgen der Länge N analysieren, wird (3.67) durch

$$c'(o) = \frac{1}{N} \sum_{k=o}^{N-1} \ln |Y(k)| \tag{3.71}$$

approximiert. Durch Überfaltung (aliasing) aufgrund der Benutzung der DFT hängt c'(o) mit den exakten Koeffizienten c(n) über

$$c'(o) = \sum_{k=-\infty}^{\infty} c(kN) \tag{3.72}$$

zusammen (GOLD & RADER, 1969). Da eine Eigenschaft des Cepstrums ist (OPPENHEIM & SCHAFER, 1975), daß

$$|c(n)| < K \left|\frac{\gamma}{n}\right|^{|n|} \quad \text{mit } |\gamma| < 1 \text{ und } K > 0 \tag{3.73}$$

ist, ist der Fehler, der dadurch entsteht, daß c'(o) durch c(o) ersetzt wird, bei genügend großem N vernachlässigbar. Die mit Hilfe von c'(o) berechneten normierten Fehler V_∞ sind in Bild 3.9 im Vergleich zu V_{min} gezeigt.

Aus (3.62) kann man noch zwei Eigenschaften der Approximation der beiden Spektren aneinander ablesen. Zum einen geht aufgrund der Fehlerbildung durch den Quotienten der Betragsspektren die Approximation gleichmäßig über den gesamten Frequenzbereich. Zum anderen gilt wegen (3.62) und (3.64)

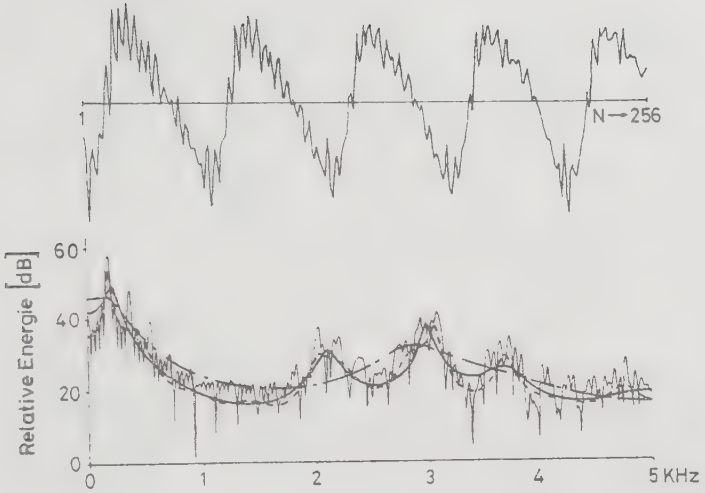


Bild 3.11 Approximation des logarithmierten Betragsspektrums der abgebildeten Probe stimmhafter Sprache durch Prädiktion mit unterschiedlicher Ordnung p.
—·—· ≙ p = 6; —— ≙ p = 13; – – – ≙ p = 25

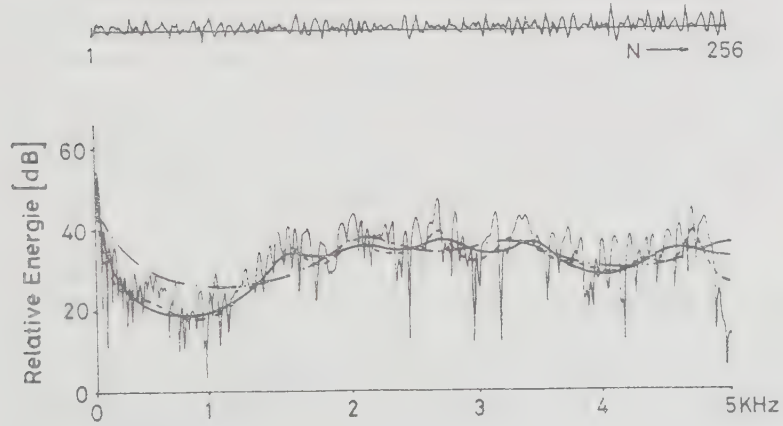


Bild 3.12 Approxiamtion des logarithmierten Betragsspektrums der abgebildeten Probe stimmloser Sprache durch Prädiktion mit unterschiedlicher Ordnung p
—·—· ≙ p = 6; —— ≙ p = 13; – – – ≙ p = 25

$$\frac{1}{N} \sum_{k=0}^{N-1} \frac{|Y(k)|^2}{|H(k)|^2} = 1 \qquad (3.74)$$

wobei man

$$E(k) = \frac{|Y(k)|^2}{|H(k)|^2} \qquad (3.75)$$

als "Augenblicksfehler" bezeichnen kann, dessen arithmetischer Mittelwert gleich eins ist. Da Werte E(k) für $|Y(k)|^2 > |H(k)|^2$ stärkeren Einfluß auf den Mittelwert haben als Werte E(k) für $|H(k)|^2 > |Y(k)|^2$ kann man erwarten, daß die Approximation in den Bereichen $|Y(k)|^2 > |H(k)|^2$ besser ist als in den Bereichen $|H(k)|^2 > |Y(k)|^2$ (MAKHOUL, 1975). Bild 3.11 und Bild 3.12 verdeutlichen die Auswirkung dieser Aussage, daß nämlich $|H(k)|$ eine gute Approximation der Einhüllenden von $|Y(k)|$ ist. Je größer nun die Ordnung p der linearen Prädiktion wird, desto mehr wird auch die Feinstruktur des Betragsspektrums nachgebildet. In der Sprachsynthese sind wir natürlich nur an einer Approximation der Einhüllenden interessiert, da die Feinstruktur durch die harmonische Anregung bzw. Rauschanregung entsprechend Bild 3.3 automatisch entsteht. Aufgrund der in den Bildern 3.9, 3.11, 3.12 gezeigten experimentellen Ergebnisse ergibt sich ein Wert von p zwischen 10 - 15 als geeignet. Für die weiteren experimentellen Untersuchungen haben wir den Wert p = 13 gewählt.

3.2.5.2 Einfluß der Fensterlänge und der Fensterform

Experimentell wurden in einer Arbeit(CHANDRA & LIN, 1974) verschiedene Parametereinflüsse, darunter auch die Fensterlänge untersucht. Es wurde jeweils der durch (3.60) definierte normierte Fehler V in Abhängigkeit der verschiedenen Parameter dargestellt. Dabei wurde allerdings der minimal mögliche Fehler V_∞ nicht mitdargestellt, der unserer Meinung nach aber erst eine genauere Beurteilung zuläßt. Bild 3.13 zeigt sowohl den Fehler V als auch V_∞ als Funktion der Fensterlänge NW. Das

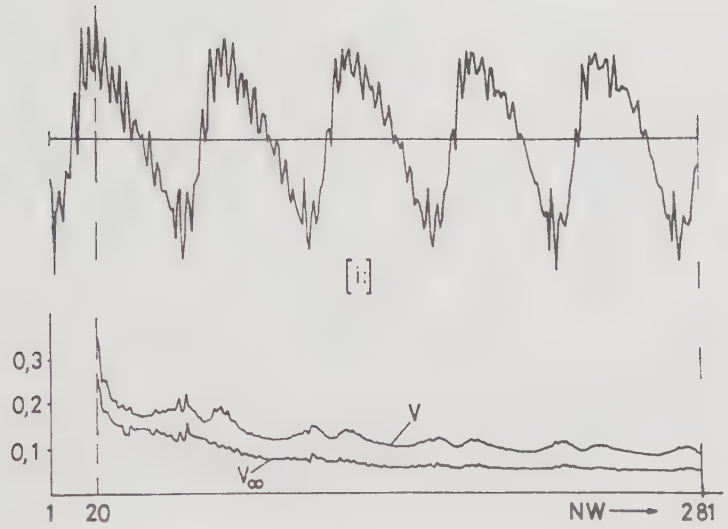


Bild 3.13 Normierter minimaler quadratischer Fehler V für p = 13 und minimal möglicher Fehler V_∞ als Funktion der Fensterlänge NW für die abgebildete Probe stimmhafter Sprache

zugrundeliegende Sprachsignal ist der schon in Bild 3.10 a dargestellte Ausschnitt aus dem Vokal [i:], der dem Wort Biene entnommen ist. Zur Verdeutlichung haben wir das Sprachsignal entsprechend über dem zugehörigen Fehler aufgetragen. Wie man erkennen kann, haben nur bei kleinen Werten von NW die jeweils in den Fensterbereich eintretenden Glottispulse einen sichtbaren Einfluß auf die beiden Fehlergrößen. Auch bei dem demselben Wort entnommenen Endvokal [ə], bei dem, wie Bild 3.14 zeigt, die Glottispulse sehr deutlich sichtbar sind, wird bei großen Fensterlängen ihr Einfluß gering. Bei stimmloser Sprache - Bild 3.15 zeigt einen Ausschnitt aus dem Zischlaut [ʃ], der dem Wort Schwelle entnommen ist - sind, wie das gleiche Bild deutlich macht, aufgrund des rauschartigen Charakters des Zeitsignals die Fehlerfunktionen ebenfalls rauschartig und wenig von der Fensterlänge abhängig. Bei allen bis jetzt vorgestellten Beispielen ist die Fensterfunktion

$$w_R(n) = \begin{cases} 1 & n_o \le n \le n_o + NW - 1 \\ 0 & \text{für alle anderen } n \end{cases} \qquad (3.76)$$

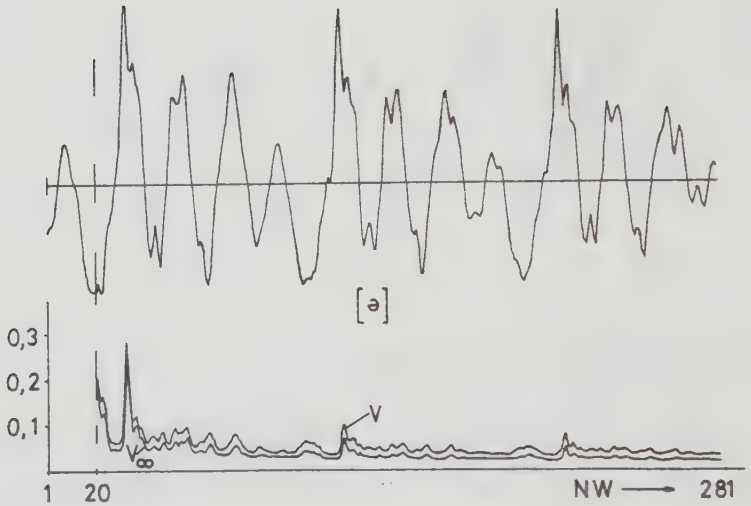


Bild 3.14 Normierter minimaler quadratischer Fehler V für p = 13 und minimal möglicher Fehler V_∞ als Funktion der Fensterlänge NW für die abgebildete Probe stimmhafter Sprache

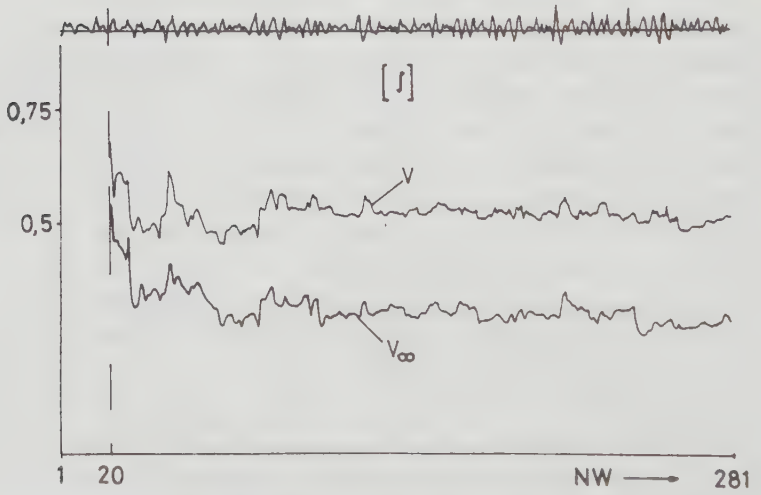


Bild 3.15 Normierter minimaler quadratischer Fehler V für p = 13 und minimal möglicher Fehler V_∞ als Funktion der Fensterlänge NW für die abgebildete Probe stimmloser Sprache

das sogenannte Rechteckfenster. Durch eine Fensterfunktion, die die am Rande des Fensters liegenden Signalwerte weniger stark gewichtet als die in der Mitte liegenden, lassen sich die in den Bildern 3.13. und 3.14 gezeigten Einflüsse der Glottispulse natürlich verringern. Eine häufig benutzte Fensterfunktion ist das Hamming-Fenster, das definiert ist durch

$$w_H(n) = \begin{cases} 0{,}54+0{,}46\ \cos[2\pi(n_o-n)/NW-1] & \text{für } n_o \leq n \leq n_o+NW-1 \\ 0 & \text{für alle anderen n} \end{cases} \qquad (3.77)$$

Bild 3.16 zeigt den Einfluß des Hamming-Fensters auf die Fehler V und V_∞ als Funktion der Fensterlänge NW für die drei

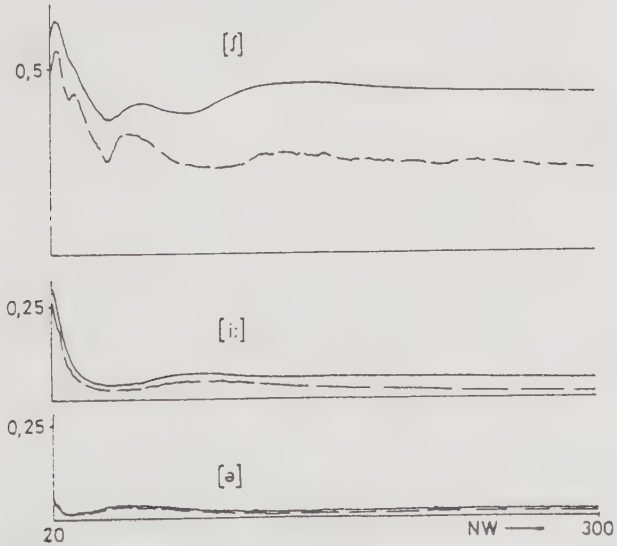


Bild 3.16 Einfluß des Hamming-Fensters auf den normierten minimalen quadratischen Fehler V und den minimal möglichen Fehler V_∞

vorgestellten Beispiele. Wie wir erkennen können, ist für NW > 150 der Einfluß der Änderungen des Zeitsignals vernachlässigbar. Für die Analyse benutzen wir deshalb und auch aus Vereinfachungsgründen beim Datentransfer von der Magnetplatte eine Fensterlänge von NW = 256.

3.2.5.3 Einfluß der Position des Fensters

Neben der Fensterlänge und der Fensterform wurde von uns auch der Einfluß der relativen Lage des Fensters im Sprachsignal untersucht. Zwar liegen aus einer von RABINER et al., 1977 gemachten Untersuchung einige Ergebnisse vor, aber auch in dieser Untersuchung wird der Fehler V nicht im Zusammenhang mit den minimal möglichen Fehler V_∞ betrachtet. Um diese Fehlergrößen in Abhängigkeit von der Lage im Sprachsignal darzustellen, werden für ein Fenster gegebener Länge sowohl der normierte minimale quadratische Fehler V, als auch der Fehler V_∞ berechnet; das Fenster wird dann um einen Abtastwert verschoben, und die beiden Fehlergrößen werden neu bestimmt. Bild 3.17

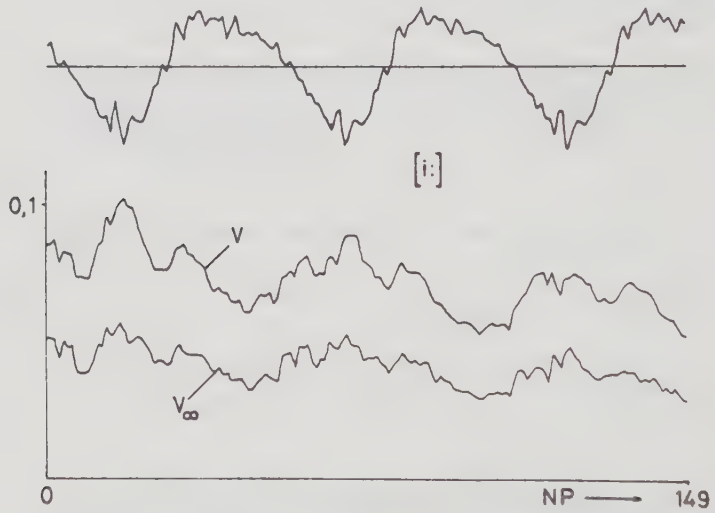


Bild 3.17 Normierter minimaler quadratischer Fehler V für p = 13 und NW = 256 und minimal möglicher Fehler V_∞ als Funktion der Verschiebung NP

zeigt die beiden Fehler V und V_∞ als Funktion der Anzahl NP der Abtastwerte, um die das Fenster von der Ausgangsposition verschoben wurde. Die Fensterlänge ist NW = 256 und die Fensterfunktion ist das Rechteckfenster (3.76). Die untersuchte Sprachprobe ist wieder der Vokal [i:] aus dem Wort Biene. Zur Verdeutlichung sind die Abtastwerte des Sprachsignals, die durch die Verschiebung des Fensters in den Bereich des Fensters kommen, darüber dargestellt. Deutlich ist der Zusammenhang

zwischen der Lage der Glottispulse und den Fehlermaxima erkennbar. Noch deutlicher wird dieser Zusammenhang in Bild 3.18,

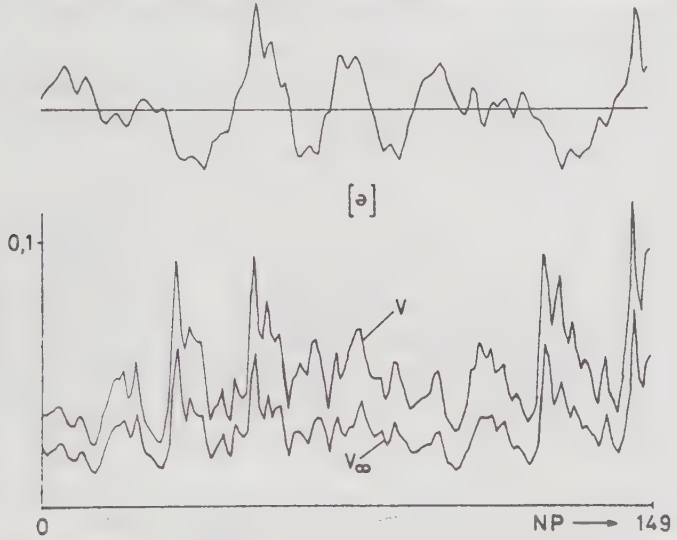


Bild 3.18 normierter minimaler quadratischer Fehler V für p = 13 und NW = 256 und minimaler möglicher Fehler V_∞ als Funktion der Verschiebung NP

in dem die gleichen Funktionen für den Vokal [ə] aus dem Wort Biene dargestellt sind. Die Verwendung eines Hamming-Fensters (3.77) verringert diese Einflüsse sehr drastisch wie Bild 3.19 verdeutlicht, in dem für beide Sprachproben sowohl der Fehler V als auch der minimal möglich Fehler V_∞ dargestellt sind. Der Vokal [i:] aus dem Wort Biene entspricht einem, der in der Untersuchung von RABINER et al. 1977 benutzten Vokale aus dem englischen Wort beet. Abweichend von unseren Ergebnissen ist bei RABINER der Einfluß des Hamming-Fensters bei mit gleicher Abtastfrequenz abgetasteten Sprache, einer vergleichbaren Prädiktorordnung von p = 14 und einer Fensterlänge NW = 200, wie Bild 3.20 zeigt, nicht so groß. Allerdings sind bei Verwendung des Rechteckfensters die Schwankungen des Fehlers V von Abtastwert zu Abtastwert auch schon erheblich größer als bei unseren Untersuchungen, bei denen aber auch die großen Unterschiede im Fehler V beim Vokal [ə] fast vollständig durch das Hamming-Fenster beseitigt werden.

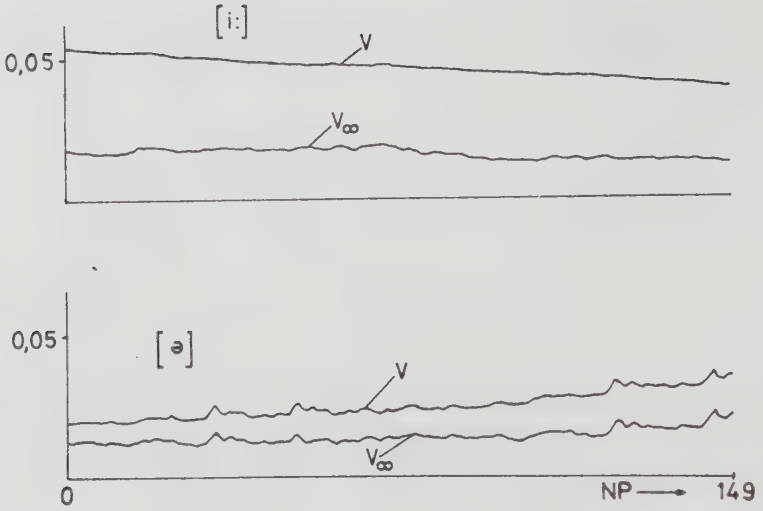


Bild 3.19 Einfluß des Hamming-Fensters auf die in Bild 3.17 und 3.18 dargestellten Fehlerfunktionen

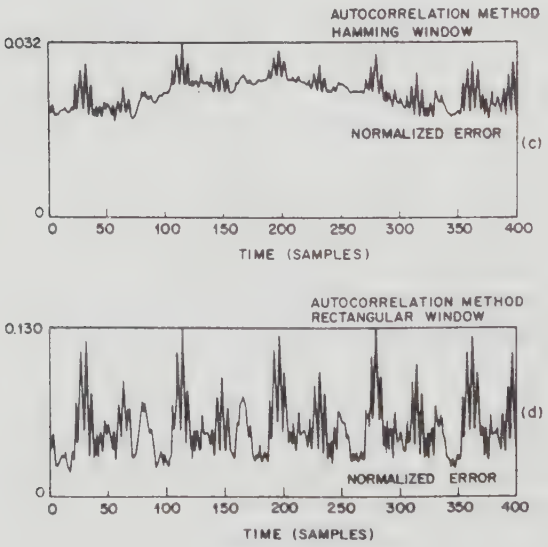


Bild 3.20

Einfluß des Hamming-Fensters auf den normierten minimalen quadratischen Fehler V (RABINER et al., 1977)

Für die Analyse des Sprachsignals scheint also nach dem vorher Gesagten eine Fensterlänge NW = 256 und die Bewertung des Fensters durch ein Hamming-Fenster geeignet. Bild 3.21 und 3.22 zeigen die den Bildern 3.11 und 3.12 entsprechenden Ergebnisse bei der Bewertung durch ein Hamming-Fenster. Weitere Ergebnisse über den Einfluß der Analyse-Parameter wollen wir im folgenden durch erste Beurteilung der synthetisch erzeugten Sprache gewinnen.

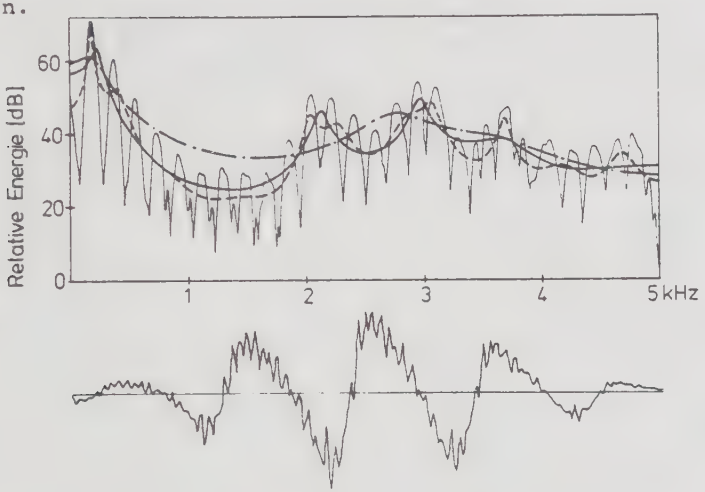


Bild 3.21 Approximation der mit einem Hamming-Fenster bewerteten Sprachprobe entsprechend dem Bild 3.11

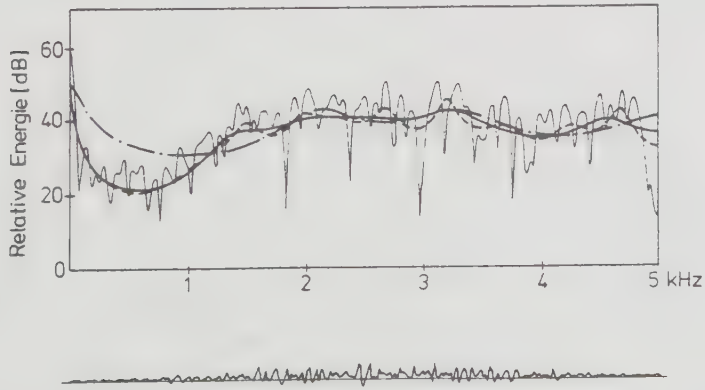


Bild 3.22 Approximation der mit einem Hamming-Fenster bewerteten Sprachprobe entsprechend dem Bild 3.12

4. Anregungsfunktion

Wie wir schon im Kapitel 3 dargestellt haben, läßt sich das digitale Leitungsmodell einerseits als Nachbildung des Vokaltraktes in einem Spracherzeugungsmodell entsprechend Bild 3.1 einsetzen, andererseits kann es auch als Prädiktionsfilter in dem in Bild 3.2 dargestellten vereinfachten Modell benutzt werden. Die Eingangssignale, die zur Sprachsynthese durch das Leitungsmodell benötigt werden, sind natürlich nicht gleich und sollen im folgenden näher untersucht werden.

4.1 Anregungsfunktion für das Vokaltraktmodell

4.1.1 Stimmhafte Anregung

In dem in Bild 3.1 gezeigten Modell zur Spracherzeugung bildet das Leitungsmodell den natürlichen Vokaltrakt nach. Das auf dieses Modell wirkende Eingangssignal sollte deswegen dem natürlichen Glottissignal möglichst ähnlich sein. Obwohl die Eigenschaften des Glottissignals und ihr Einfluß auf die Spracherzeugung schon recht lange Gegenstand der Forschung waren - z.B. HELMHOLTZ, 1877 und WHEATSTONE, 1879 - lagen exakte Aussagen lange nicht vor. Dabei spielten sowohl die für direkte Messungen ziemlich unzugängliche Lage der Stimmbänder, als auch ungenügende Kenntnis über den Einfluß des Vokaltraktes eine Rolle. Die am häufigsten benutzten Verfahren zur Ermittlung der Zeitfunktion des Glottissignals lassen sich in zwei Kategorien einteilen. Zur ersten Kategorie gehört das Filmen der Stimmbänder. Diese Methode wurde erstmals von FARNSWORTH, 1940 angewandt. Aus den mit 4000 Bildern pro Sekunde gemachten Filmaufnahmen läst sich die Fläche A_G zwischen den Stimmbändern als Funktion der Zeit bestimmen. Diese Flächenfunktion $A_G(t)$ benutzte FLANAGAN, 1958, um den Volumenfluß $q_G(t)$ zu ermitteln. Weitere Arbeiten in dieser Kategorie wurden von BEROUTI et al., 1977 und NADAL-SURIS & CHILDERS, 1977 durchgeführt. bzw. werden von diesen zitiert. Die zweite Kategorie besteht aus Verfahren, bei denen das Glottissignal aus dem akustischen Signal an den Lippen ermittelt wird. Die Filterwirkung des Vokaltrakts wird dabei durch ein inverses Filter,

d.h. ein Filter, das eine Übertragungsfunktion hat, die invers zur Übertragungsfunktion des Vokaltraktes ist, kompensiert. Die Übertragungsfunktion wird dabei natürlich nur angenähert. Die aufgezeichnete akustische Größe kann dabei sowohl der Schalldruck (MILLER, 1959; MATHEWS et al., 1961) als auch der Volumenfluß (ROTHENBERG, 1973; ROTHENBERG & ZAHORIAN, 1977) sein. Ein weiteres Verfahren in dieser Kategorie, das ohne die inverse Filterung auskommt, wurde von SONDHI, 1975 vorgestellt. Dabei wird der mit einer über der Länge relativ konstanten Querschnittsflächenfunktion A(x) gebildete sogenannte Schwa-Laut [ə] in ein reflexionsfrei abgeschlossenes Rohr gleicher Querschnittsfläche artikuliert. Mit einem Mikrofon in der Wand des Rohres kann nun, da die Anordnung im Idealfall nur eine zeitliche Verzögerung mit sich bringt, direkt das Glottissignal aufgezeichnet werden.

Die Funktion der Stimmbänder läßt sich natürlich auch durch entsprechende Modelle beschreiben. Ausgehend von den physiologischen Gegebenheiten wurde zunächst von FLANAGAN & LANDGRAF, 1969 ein mechanisches Modell, ein mit einer Differentialgleichung 2. Ordnung beschreibbarer mechanischer Oszillator, entwickelt. Da dieses Modell sich als zu grob erwies, wurde es von ISHIZAKA & FLANAGAN, 1972 verbessert. Die Funktion der Stimmbänder wird dabei durch zwei über eine Feder miteinander verkoppelte mechanische Oszillatoren nachgebildet. Dieses Modell ist ein Teil eines durch physiologische Parameter gesteuerten Rechnermodells der Spracherzeugung (FLANAGAN et al., 1975). Eine Schaltung, die dieses Modell als in Echtzeit arbeitende Quelle zur Sprachsynthese realisiert, wurde von CARCAUD et al., 1976 vorgestellt. Eine zweite mögliche Art, die Modellbildung durchzuführen, wurde von ROTHENBURG et al., 1974 praktiziert. Dabei steht nicht wie bei FLANAGAN die physiologische Nachbildung im Vordergrund. Es wird vielmehr der funktionale Zusammenhang zwischen drei physiologisch ermittelten Variablen und dem resultierenden Glottissignal durch ein System nachgebildet, das nicht notwendigerweise den physiologischen Vorgängen entspricht. Als Vergleichsgröße für die Güte des Ausgangssignals dienten dabei die von ROTHENBERG, 1973 durch inverse Filterung gewonnenen Glottissignale.

4.1.2 Stimmlose Anregung

Bei der stimmlosen Anregung ist der Kenntnisstand weit geringer als bei der stimmhaften Anregung. Ausgehend von Untersuchungen von MEYER-EPPLER, 1953 erweiterten FLANAGAN & CHERRY, 1969 und FLANAGAN & ISHIZAKA, 1976 das schon oben erwähnte Rechnermodell der Spracherzeugung um die stimmlose Anregung. Dabei geht man davon aus, daß durch Verengungen im Vokaltrakt die normalerweise laminare Strömung turbulent wird. Die dabei auftretenden unregelmäßigen Druckschwankungen werden dabei als Geräusch hörbar. Jedem Leitungsstück mit Index n in dem Rechnermodell von FLANAGAN wird dabei eine Rauschschalldruckquelle P_n mit einem zugehörigen Quellenwiderstand R_n zugeordnet. Sowohl die Amplitude der Rauschquelle als auch der Wert des Widerstandes R_n sind dabei Funktionen der Querschnittsfläche A_n des Leitungsstückes und des Volumenflußes q_n in dem entsprechenden Leitungsstück n. Die Amplitude ist dabei

$$P_n \sim \frac{q_n^2}{A_n} \tag{4.1}$$

Der Widerstand ist in einer ersten Näherung

$$R_n \sim \frac{|q_n|}{A_n^2} \tag{4.2}$$

Diese beiden Größen werden in dem Rechnermodell für jeden Abtastwert neu berechnet, so daß sich die stimmlose Anregung an den Verengungen des Vokaltraktes automatisch einstellt.

4.2 Anregungsfunktion für das Prädiktionsfilter

4.2.1 Stimmhafte Anregung

Um für das in Bild 3.2 dargestellte Spracherzeugungsmodell die lineare Prädiktion anwenden zu können, ist, wie wir im Kapitel 3.2.1 gesehen haben, eine Voraussetzung, daß der Betrag des

Spektrums des Eingangssignals konstant ist. Aus diesem Grunde wurde zunächst häufig die stimmhafte Anregung als Pulsquelle, die mit der Periode der Sprachgrundfrequenz Einheitsimpulse abgibt, realisiert. (ATAL & HANAUER, 1971; MARKEL & GRAY, 1974). Dadurch, daß die Energie der Anregung in diesem Falle nur durch den einen von Null verschiedenen Abtastwert des Einheitsimpulses aufgebracht wird, klingt die so erzeugte Sprache häufig "rauher" als die natürliche Sprache. Um diese Rauhigkeit - in der englischen Literatur "buzziness" genannt - zu beseitigen, kann man die noch vorhandenen Freiheitsgrade bei der Zeitfunktion der Eingangsgröße ausnutzen. Der starke Impulscharakter der Anregungsfunktion kann durch Verwendung von Zeitfunktionen, die mehrere von Null verschiedene Abtastwerte umfassen, verringert werden. In einer Arbeit von SAMBUR et al., 1978 wurden dabei von ROSENBERG, 1971 ermittelte Zeitfunktionen als Anregungsfunktion für ein Prädiktionsfilter benutzt. Dabei ergab sich, daß die Kurvenform keine entscheidende Rolle bei der Qualitätsbeurteilung der erzeugten Sprache spielt, sondern nur die Zeitdauer der ansteigenden bzw. abfallenden Flanke des Signals entscheidend ist. Als einfachsten Fall einer solchen Zeitfunktion schlugen sie eine Dreiecksfunktion vor. Zu ähnlichen Ergebnissen kamen KUREMATSU et al., 1979. Bei all diesen Zeitfunktionen weicht man natürlich von der Forderung nach einem konstanten Betragsspektrum ab, denn zu hohen Frequenzen hin sinkt der Betrag merklich ab. Diese Eigenschaft hat ein von WONG & MARKEL, 1978 vorgestelltes Anregungssignal nicht. Es besitzt außerdem noch die nützliche Eigenschaft, daß sein arithmetischer Mittelwert Null ist. Außerdem haben sich WONG & MARKEL bemüht, die Phase des natürlichen Glottissignals in dem synthetischen Anregungssignal zu erhalten, da diese Maßnahme die Qualität bei Darbietung über Kopfhörer verbessern soll.

4.2.2 Stimmlose Anregung

Um für die stimmlose Anregung die Forderung nach einem konstanten Betrag des Spektrums zu erfüllen, wird üblicherweise eine Folge unkorrelierter Abtastwerte (weißes Rauschen) als Eingangssignal benutzt. In den meisten bisherigen Anwendungen

wird eine der beiden Anregungsfunktionen mit Hilfe eines Umschalters als Eingangssignal auf das Prädiktionsfilter gegeben. Dabei kommt dann der Entscheidung, ob die Sprache stimmlos oder stimmhaft ist, große Bedeutung zu. Mit einer gemischten Anregungsfunktion könnten sich durch Fehler an den Übergangsstellen auftretende Qualitätseinbußen vermeiden lassen. MAKHOUL et al., 1978 stellten ein Modell einer solchen Anregungsquelle vor, die den Grad der Stimmhaftigkeit bzw. Stimmlosigkeit frequenzselektiv einstellt.

4.3 Realisierung des Anregungsgenerators

Bild 4.1 zeigt den prinzipiellen Aufbau des von FRIEDERICH, 1979 in einer Diplomarbeit realisierten Anregungsgenerators. Um mit einer Vielzahl von stimmhaften Anregungsfunktionen später Versuche machen zu können, wurde ein 256 x 14 bit RAM als Speicher für die Abtastwerte des Zeitsignals der stimmhaften Anregung gewählt. In diesen Speicher können beliebige im Laborrechner HP 21 MX erzeugte Zeitfunktionen eingeschrie-

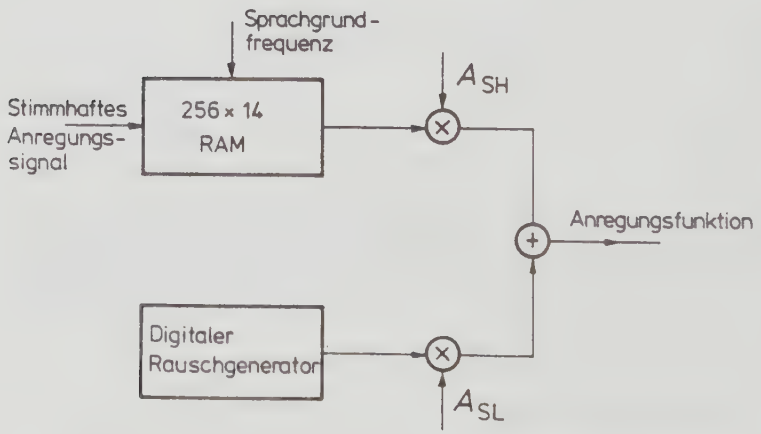


Bild 4.1 Blockschaltbild des Anregungsgenerators

ben werden. Bei der Spracherzeugung werden dann entsprechend dem vom Rechner übergebenen Wert der Periode T_{sp} der Sprachgrundfrequenz, die zugehörige Anzahl von Abtastwerten ausgegeben. Die Änderung der Zeitfunktion während der Sprachsynthese ist z.Zt. nicht möglich, läßt sich aber leicht durch eine entsprechende Erweiterung erreichen. Die Anregungsfunktion setzt sich, wie Bild 4.1 verdeutlicht, aus der jeweils mit den Amplitudenfaktoren A_{SH} bzw. A_{SL} bewerteten Summe der beiden Anregungssignale zusammen. Die Zahlenwerte von A_{SH}, A_{SL} und T_{sp} sind jeweils mit 8 bit Wortlänge quantisiert und werden dem Generator vom Laborrechner übergeben. Diese Übergabe geschieht üblicherweise im Zeitraum zwischen dem Auslesen des letzten und des ersten Abtastwertes des stimmhaften Anregungssignals. Die Parameter können sich somit maximal im Takt der Sprachgrundfrequenz ändern.

4.4 Gewinnung der Anregungsparameter aus natürlicher Sprache

Um nun mit dem in Bild 4.1 dargestellten Anregungsgenerator und dem Leitungsmodell entsprechend dem Modell der Spracherzeugung in Bild 3.2 Sprache erzeugen zu können, benötigen wir Werte für die Periode T_{sp} der Sprachgrundfrequenz und die Amplitudenfaktoren A_{SH} bzw. A_{SL}.

4.4.1 Sprachgrundfrequenz

Neben der Bedeutung bei der natürlichen Erzeugung von Sprache ist die Bestimmung der Sprachgrundfrequenz ein wesentlicher Bestandteil in Systemen der Sprechererkennung und wird in fast allen Sprach-Analyse-Synthese Systemen benötigt. Aus diesem Grunde sind eine große Vielfalt von Algorithmen zur Bestimmung der Sprachgrundfrequenz vorgeschlagen worden (zahlreiche Literaturhinweise finden sich in FLANAGAN, 1972 und RABINER & SCHAFER, 1978). Bei der Auswahl eines geeigneten Algorithmus konnten wir auf zwei Untersuchungen zurückgreifen, die objektiv bzw. subjektiv die Ergebnisse von sieben Verfahren miteinander verglichen haben (RABINER et al., 1976; McGONEGAL et al., 1977). Bei der objektiven Untersuchung wurden die Ergebnisse mit denen eines halbautomatischen, interaktiven Sprachgrundfrequenzer-

mittlungsverfahrens (SAPD) (McGONEGAL et al., 1975) verglichen. Für den von uns wichtigen subjektiven Vergleich wurde mit Hilfe der linearen Prädiktion synthetische Sprache erzeugt, wobei die Daten über die Sprachgrundfrequenz von einem der zu untersuchenden Verfahren geliefert wurden oder mit dem schon erwähnten SAPD ermittelt wurden. Als Vergleich stand jeweils auch die natürliche Sprachprobe zur Verfügung. Diese neun vorhandenen Sprachproben wurden von den Versuchspersonen mit Hilfe einer rechnergesteuerten Sortiermatrix (COKER & PRUZANSKY, 1970) entsprechend ihrer Rangfolge geordnet. Das für den Test verwandte Sprachmaterial umfaßte 4 verschiedene Sätze, die von 6 verschiedenen Sprechern unter 3 verschiedenen Aufnahmebedingungen aufgenommen worden waren. Neben der Rangfolge der Verfahren zur Ermittlung der Sprachgrundfrequenz, die bei Mittelung über die 72 verschiedenen Bedingungen ermittelt wurde, wurde auch der Einfluß der 3 oben genannten äußeren Parameter auf die Rangfolge der Verfahren untersucht. Für unsere Anwendung sind dabei besonders die Abhängigkeit von den benutzten Sätzen und von den verschiedenen Sprechern interessant. In beiden Fällen ist der Einfluß dieser beiden äußeren Parameter bei dem AMDF (Average Magnitude Difference Function) Verfahren (ROSS et al., 1974) am geringsten. Bei der über alle Parameter gemittelten Rangfolge ergibt sich bei einer Rangskala von 1 bis 9 folgendes Bild:

	Mittlerer Rang	Standardabweichung
natürliche Sprache	8,97	0,25
SAPD	6,46	1,55
LPC	5,16	2,20
AMDF	5,03	1,75
PPROC	4,56	1,68

(nach McGONEGAL et al., 1977).

Bei den eigentlich untersuchten Verfahren liegen also ein von ATAL entwickeltes, aber unveröffentlichtes LPC Verfahren und das schon erwähnte AMDF Verfahren dicht beieinander an der Spitze. Das nächste Verfahren PPROC (Parallel PROCessing) (GOLD & RABINER, 1969) fällt schon deutlich ab. (Die restlichen 4 schlechter beurteilten Verfahren sind wegen der Übersicht-

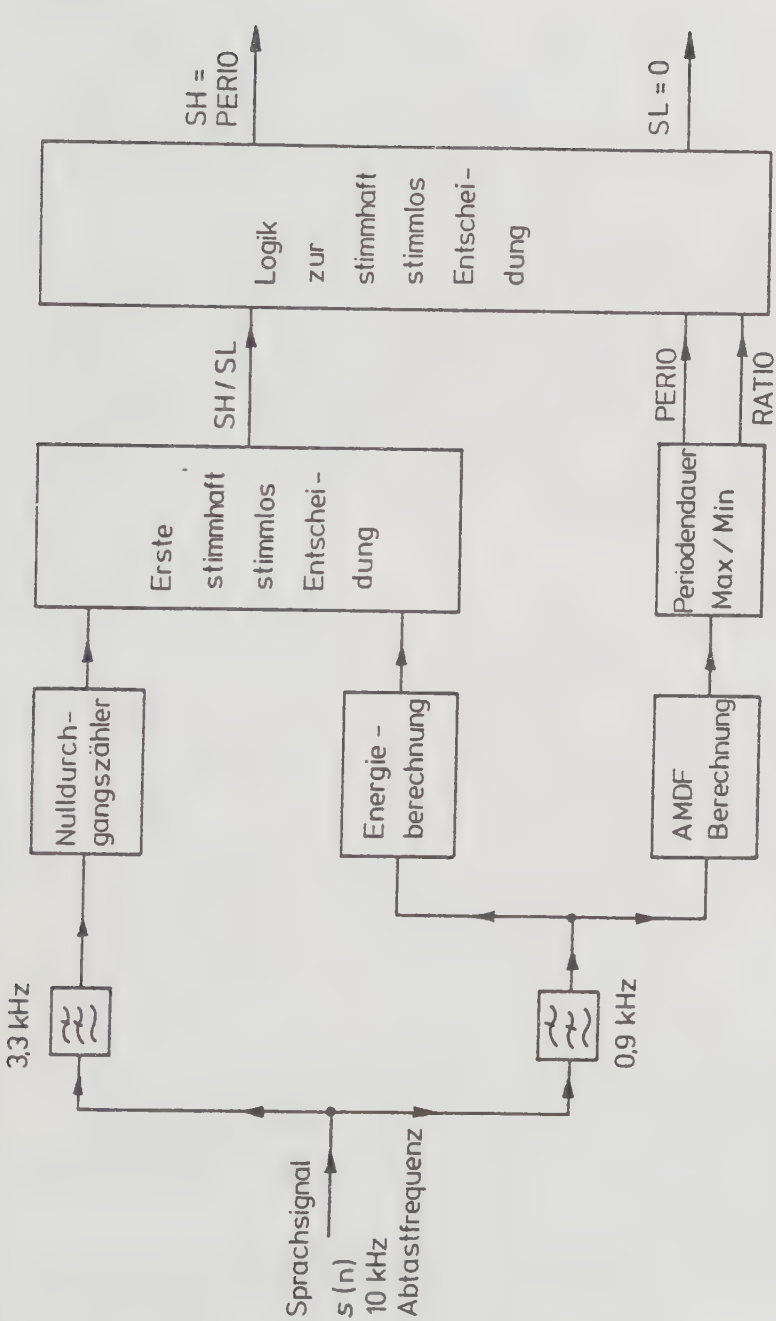


Bild 4.2 Blockschaltbild des AMDF Verfahrens

lichkeit nicht aufgeführt.

Im Rahmen einer Studienarbeit (SUCKEL, 1979) wurde das AMDF Verfahren als FORTRAN Programm auf dem Laborrechner implementiert. Bild 4.2 zeigt als Blockschaltbild die Arbeitsweise des Verfahrens. Zunächst wird das mit 10 kHz abgetastete Sprachsignal mit dem im Kapitel 3 beschriebenen nichtrekursiven Tiefpaß mit linearer Phase gefiltert. Dabei ist die Grenzfrequenz einmal 3,3 kHz und zum anderen 0,9 kHz. Beim ersten gefilterten Signal werden die Nulldurchgänge gezählt und zusammen mit der Energie, die vom zweiten Signal berechnet wird, wird eine erste stimmhaft/stimmlos Entscheidung getroffen. Aus dem auf 900 Hz begrenzten Signal wird die gemittelte Betragsdifferenzfunktion (AMDF)

$$a(n) = \frac{1}{N} \sum_{i=1}^{N} \left| s(i) - s(i-n) \right| \qquad n = 0,1,2,\ldots,M \qquad (4.3)$$

gebildet. Diese Funktion ist für n = 0 Null und für alle n, die einem Vielfachen der Periodendauer T_{sp} der Sprachgrundfrequenz entsprechen, hat sie ein starkes Minimum. Bild 4.3a,b verdeutlicht diese Eigenschaft der Funktion a(n) an zwei Beispielen. Das zugehörige Sprachsignal s(n) ist jeweils darüber dargestellt. Bei stimmloser Sprache treten, wie Bild 4.3c zeigt, diese Eigenschaften der AMDF nicht auf. Die aus dieser Funktion bestimmte Periode PERIO der Sprachgrundfrequenz und das Verhältnis zwischen dem Maximum und Minimum der Funktion werden in einem zweiten Entscheidungsschritt zusammen mit der ersten stimmhaft/stimmlos Entscheidung verarbeitet. Bei dieser Entscheidung werden auch die beiden vorhergegangenen Entscheidungen gewichtet berücksichtigt. Das Ergebnis für das jeweils betrachtete Analysefenster ist dann bei der Entscheidung stimmhaft die Periode T_{sp} der Sprachgrundfrequenz. Bei der Entscheidung stimmlos wird eine Null ausgegeben. Änderungen der Periode T_{sp} können in unserem Fall aufgrund der Abtastfrequenz von 10 kHz in Schritten von 100 µsec berücksichtigt werden. Bild 4.4 zeigt das Ergebnis der Analyse des Wortes "Ochse". Dabei

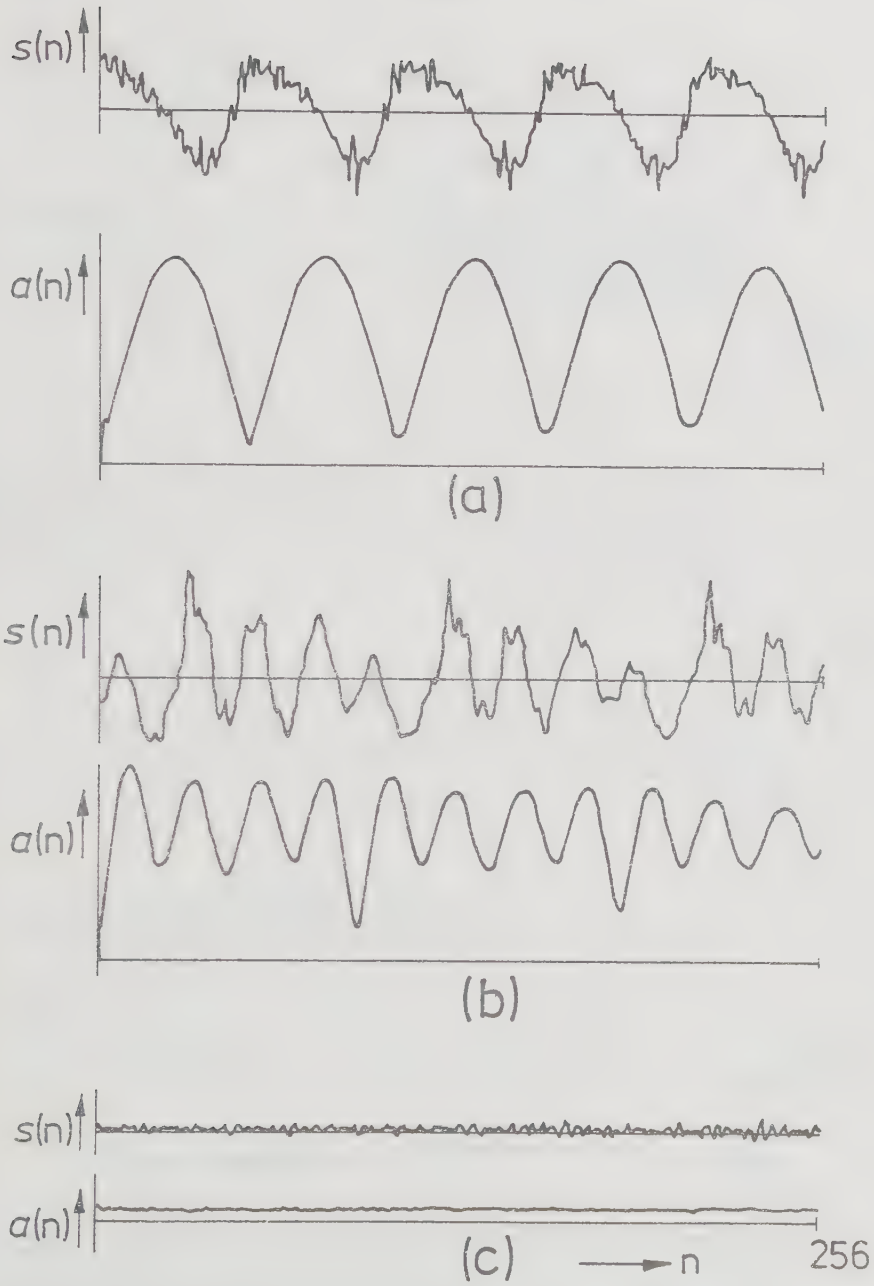


Bild 4.3 Betragsdifferenzfunktionen a(n) für stimmhafte Sprache (a), (b) und stimmlose Sprache (c)

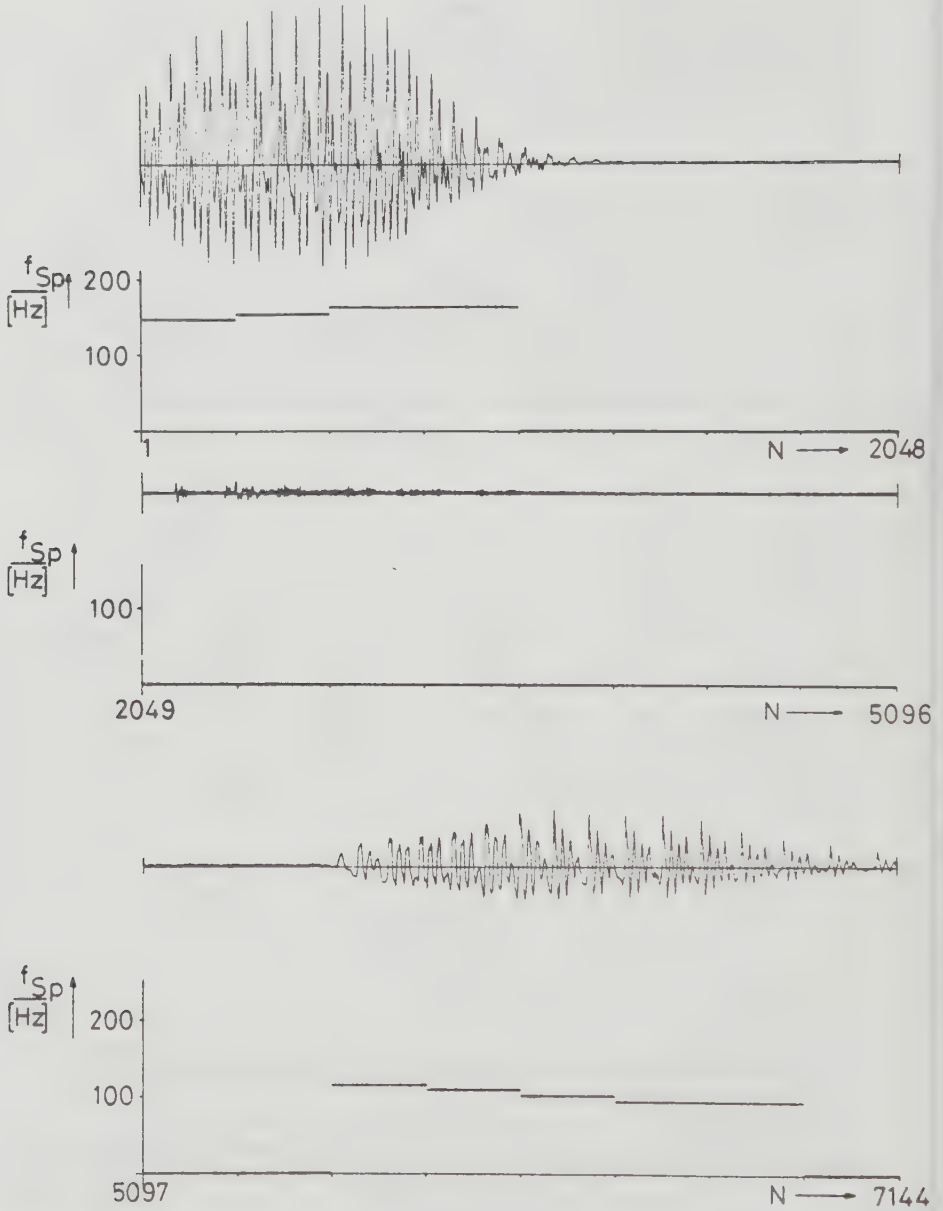


Bild 4.4 Zeitsignal des Wortes "Ochse" mit zugehöriger durch AMDF Verfahren ermittelter Sprachgrundfrequenz

ist jeweils über der Darstellung der Sprachgrundfrequenz f_{sp} das zugehörige Sprachsignal gezeigt. Die Länge des Analysefensters beträgt dabei 256 Abtastwerte (25,6 msec).

4.4.2 Amplitude

Bei der der linearen Prädiktion zugrundeliegenden Übertragungsfunktion

$$H(z) = \frac{G}{1 + \sum_{k=1}^{p} \beta_k z^{-k}}$$

dient, wie wir schon in Kapitel 3.2.1 dargestellt haben, der Verstärkungsfaktor G dazu, die Amplitude des Eingangssignals so einzustellen, daß die Energie des erzeugten Signals gleich der des ursprünglichen Signals ist. Aus (3.3 1) erkennt man, daß eine recht einfache Möglichkeit, dieses Ziel zu erreichen, ist, die Energie des mit dem Verstärkungsfaktor G bewerteten Eingangssignals in einem der Analyse entsprechenden Fenster der Länge N gleich dem minimalen Fehler D zu machen (MARKEL & GRAY, 1974). Für ein beliebiges Anregungssignal x(n) muß also

$$D = G^2 \sum_{n=1}^{N} x^2(n) \qquad (4.4)$$

gelten. Für den speziellen Fall der stimmhaften Anregung mit Einheitsimpulsen, die im Abstand der Periode T_{sp} der Sprachgrundfrequenz aufeinander folgen, ergibt sich

$$G_{SH} = \sqrt{D \frac{T_{sp}}{T\,N}} \qquad (4.5)$$

Normiert man andere mögliche stimmhafte Anregungssignale so, daß ihre Energie gleich der des Einheitsimpulses ist, läßt sich der errechnete Verstärkungsfaktor (4.5) auch bei ihnen benutzen. Bei stimmloser Anregung durch "weißes Rauschen" ergibt sich aufgrund der Energie eines solchen Signals

$$G_{SL} = \sqrt{\frac{3\ D}{N}} \tag{4.6}$$

Bei dem von uns realisierten Leitungsmodell tritt, wie wir aus (3.45) und (3.48) sehen können, im Zähler der Übertragungsfunktion H(z) der konstante Faktor

$$K = \prod_{i=1}^{p} (1 - \alpha_i) \tag{4.7}$$

auf. Um unser Leitungsmodell als Prädiktionsfilter zu benutzen, muß der jeweils auftretende Verstärkungsfaktor G durch den Faktor K dividiert werden, damit wir identische Übertragungsfunktionen erhalten.

5. Rechnersimulation

Sowohl um Größen für den schaltungsmäßigen Aufbau des Leitungsmodells, z.B. Koeffizientenwortlänge, festlegen zu können, als auch zur intensiven Analyse des Modells, wurde eine Simulation des Leitungsmodells auf dem Laborrechner implementiert. Dabei wurde insbesondere auf eine exakte Nachbildung der in der realisierten Schaltung auftretenden arithmetischen Operationen Addition und Multiplikation geachtet. Zu diesem Zweck wurden in der Programmiersprache ASSEMBLER Unterprogramme geschrieben, die sowohl den Adaptor in allgemeiner Lage als auch den Abschluß und die Einspeisung nachbilden. Zusammen mit einem übergeordneten FORTRAN Programm, das die Übergabe der Parameter und Daten übernimmt und eventuelle Meldungen auf der Konsole ausgibt, ergibt sich somit ein Unterprogramm IWDFI, das das Leitungsmodell simuliert, wobei die Anzahl der Leitungsstücke beliebig ist.

Dieses Unterprogramm wird zum einen in dem Programm IWDIM benutzt, das die Impulsantwort des Leitungsmodells und über die diskrete Fourier Transformation das zugehörige Spektrum berechnet. Zum anderen wird es vom Programm IWDSY benötigt, das mit Hilfe der ermittelten Parameter Sprache synthetisiert. (Flußdiagramme im Anhang)

5.1 Koeffizientenwortlänge

Die vom Programm LPCAN ermittelten Koeffizienten k_i, die, wie wir gezeigt haben, den von unserem Leitungsmodell zur Sprachsynthese benötigten Adaptorkoeffizienten α_i entsprechen, werden von diesem in Gleitkommadarstellung auf der Magnetplatte des Laborrechners abgelegt. Für die Realisierung kommt aus Gründen der Einfachheit der Schaltung nur eine Festkommadarstellung in Frage. Dabei ist die Entscheidung über die notwendige Wortlänge ein wichtiger Schritt im Verlauf der Realisierung. Mit Hilfe der Simulation des Leitungsmodells lassen sich hier wichtige Entscheidungshilfen gewinnen.

Den Einfluß der Koeffizientenwortlänge auf das Betragsspektrum zeigt Bild 5.1. Es stellt jeweils den logarithmierten Betrag

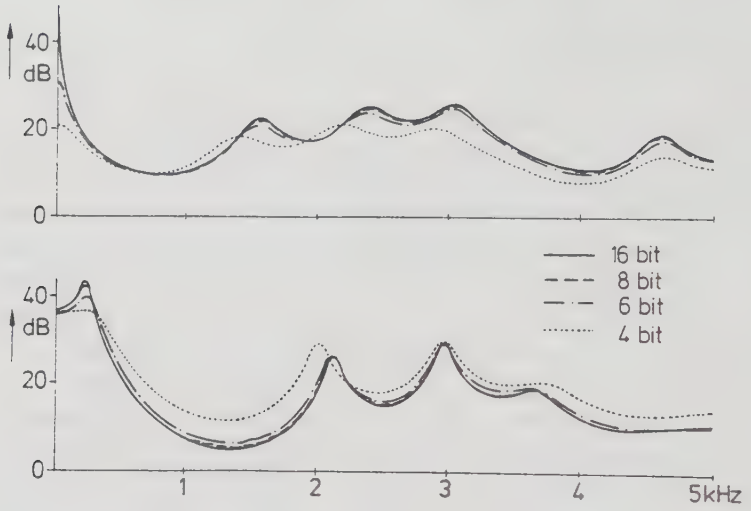


Bild 5.1 Einfluß der Koeffizientenwortlänge auf den logarithmierten Betrag des Spektrums der Impulsantwort des Prädiktionsfilters für zwei Beispiele stimmhafter und stimmloser Sprache

des Spektrums der ersten 1024 Abtastwerte der mit dem Programm IWDIM berechneten Impulsantwort h(i) dar. Der variable Parameter ist die Wortlänge der Koeffizienten. Dabei ist die Signalwortlänge entsprechend der Festkommadarstellung des Laborrechners 16 bit, während die Koeffizientenwortlänge zwischen 1 und 16 bit frei wählbar ist. Die Verkürzung der Koeffizientenwortlänge geschieht dabei durch Betragsschneiden. Diese Methode stellt sicher, daß sowohl bei positiven als auch bei negativen Zahlen bei Abschneiden der nicht gewünschten Stellen der Betrag der Zahl immer kleiner oder gleich dem Betrag der Zahl mit voller Wortlänge ist.

Wie man an den beiden Beispielen (5.1 a stimmloser Laut, 5.1 b stimmhafter Laut) sieht, ergeben sich zwischen einer Koeffizientenwortlänge von 16 und 8 bit nur sehr geringe Unterschiede, die sich als Abweichungen in der Amplitude, aber nicht in Verschiebungen der Minima und Maxima (Formanten) bemerkbar machen. Diese Art der Abweichung tritt sichtbar erst bei Koeffizientenwortlängen von 6 bzw. 4 bit auf. Als Ver-

gleich für die Güte der Festkommarealisierung wurde mit einem weiteren Rechnerprogramm, in dem sowohl die Signale als auch die Koeffizienten durch Gleitkommazahlen dargestellt werden, ebenfalls die Impulsantwort und das Betragsspektrum berechnet. Bei einer graphischen Darstellung der damit errechneten Spektren liegen die Abweichungen innerhalb der Strichstärke der Zeichnung. Die Abweichungen zwischen dem aus der mit Gleitkommazahlen berechneten Impulsantwort gewonnenen logarithmierten Betragsspektrum $X_{GK}(i)$ und dem entsprechenden Spektrum $X_{16}(i)$, bei dem die Impulsantwort mit Festkommazahlen und 16 bit Koeffizientenwortlänge berechnet wurde, ermittelt das Programm SPEDI. Es berechnet sowohl die mittlere Abweichung

$$F = \frac{1}{N} \sum_{i=1}^{N} \left| X_{GK}(i) - X_{16}(i) \right| \qquad (5.1)$$

als auch die maximal auftretende Differenz

$$DMAX = \max \left| X_{GK}(i) - X_{16}(i) \right| \qquad i = 1,2,\ldots,N \qquad (5.2)$$

zwischen zwei diskreten Spektren.

Die Anzahl der Frequenzwerte N beträgt auf Grund der benutzten ersten 1024 Abtastwerte der Impulsantwort N = 512. Für willkürlich aus zehn verschiedenen Lauten entnommene Koeffizientensätze ergibt sich für die mittlere Abweichung zwischen $X_{GK}(i)$ und $X_{16}(i)$

$$F_{16} = 0{,}046 \text{ dB} \qquad (5.3a)$$

mit einer Standardabweichung von

$$\sigma = 0{,}019 \text{ dB} \qquad (5.3b)$$

und für die maximale Abweichung

$$DMAX_{16} = 0{,}85 \text{ dB} \qquad (5.3c)$$

mit einer Standardabweichung von

$$\sigma = 0{,}5 \text{ dB} \tag{5.3d}$$

Erfaßt man die, wie Bild 5.1 verdeutlicht, in dem Maßstab gerade sichtbaren Unterschiede zwischen der $X_{GK}(i)$ und $X_8(i)$, so ergibt sich für die oben erwähnten zehn Koeffizientensätze

$$F_8 = 0{,}236 \text{ dB} \tag{5.4a}$$

mit

$$\sigma = 0{,}056 \text{ dB} \tag{5.4b}$$

und

$$DMAX_8 = 1{,}94 \text{ dB} \tag{5.4c}$$

mit

$$\sigma = 0{,}65 \text{ dB} \tag{5.4d}$$

Aufgrund der endlichen Anzahl der zur Berechnung der diskreten Fourier Transformation benutzten Abtastwerte der Impulsantwort h(i), ist das Spektrum natürlich nicht identisch mit der Übertragungsfunktion des Systems. Es ist vielmehr das Spektrum des periodisch fortgesetzten Ausschnittes der Impulsantwort. Da die Abtastwerte der 10 betrachteten Impulsantworten aber selbst bei der Berechnung mit Gleitkommazahlen für $i < 2N$ dem Betrage nach kleiner als der millionste Teil des maximalen Abtastwertes werden, sind die Abweichungen sehr gering. Um genauere Aussagen über die Abweichungen machen zu können, wurde das Programm TRNSF erstellt, das den logarithmierten Betrag der Übertragungsfunktion H(z) für $z = e^{j\omega T}$ mit $0 < \omega < \frac{\pi}{T}$, wobei T die Betriebsperiode des digitalen Filters ist, berechnet. Mit dem schon erwähnten Programm SPEDI lassen sich die Abweichungen ermitteln. Die den schon oben genannten Werten entsprechenden Größen für die Differenzen zwischen den logarithmierten Beträgen der Übertragungsfunktion H(z) und des Spektrums $X_{GK}(i)$ sind

$$F_H = 0{,}065 \text{ dB} \tag{5.5a}$$

mit

$$\sigma = 0{,}007 \text{ dB} \tag{5.5b}$$

und

$$DMAX = 0{,}302 \text{ dB} \tag{5.5c}$$

mit

$$\sigma = 0{,}103 \text{ dB} \tag{5.5d}$$

Aufgrund dieser geringen Unterschiede halten wir es für gerechtfertigt, die diskrete Fourier Transformation der ersten 1024 Punkte der Impulsantwort als gute Approximation der Übertragungsfunktion zu betrachten.

Sowohl die in Bild 5.1 sichtbaren Differenzen und Verschiebungen im Spektrum bei einer Realisierung mit 6 bit Koeffizientenlänge im Vergleich zur Gleitkommarealisierung als auch die Werte für

$$F_6 = 0{,}935 \text{ dB} \tag{5.6a}$$

mit

$$\sigma = 0{,}168 \text{ dB} \tag{5.6b}$$

und

$$DMAX_6 = 6{,}114 \text{ dB} \tag{5.6c}$$

mit

$$\sigma = 1{,}79 \text{ dB} \tag{5.6d}$$

rechtfertigen die Entscheidung für eine Realisierung mit 8 bit Koeffizientenwortlänge.

5.2 Aussteuerbarkeit

Eine weitere wichtige Größe bei der Festkommarealisierung von digitalen Systemen ist die Aussteuerbarkeit, d.h. die Antwort auf die Frage, wie groß darf die Eingangsgröße im Verhältnis zum maximal möglichen Signal sein, ohne daß die Zustandsgrößen den zugelassenen Zahlenwertebereich überschreiten. Bei der Simulation auf dem Rechner wird dazu von dem Unterprogramm IWDFI bei jeder Berechnung eines Ausgangswertes der Zustandsvariablenspeicher nach dem Maximalwert abgesucht, mit dem bisherigen größten Wert verglichen und gegebenenfalls als neues Maximum eingesetzt. Nach Abarbeitung aller Eingangswerte wird die größte aufgetretene Zustandsgröße ausgegeben. Da Überläufe beim Addieren zweier Festkommazahlen hierdurch nicht erfaßt werden, wird bei jeder in den Adaptoren durchgeführten Addition nachgeprüft, ob ein Überlauf auftritt. Ist dies der Fall, erscheint eine dementsprechende Meldung auf der Konsole.

Wie wir in Kapitel 2.4 gezeigt haben, hat das digitale Leitungsmodell mit dem Koeffizientensatz $\{\alpha_G, \alpha_1, \alpha_2, \ldots, \alpha_N, \alpha_S\}$ die gleiche Übertragungsfunktion wie mit dem Koeffizientensatz $\{\alpha_S, \alpha_N, \alpha_{N-1}, \ldots, \alpha_1, \alpha_G\}$. Für den speziellen Fall des Prädiktionsfilters, bei dem die Koeffizienten α_G und α_S entweder Null und Eins oder Eins und Null gesetzt werden, ergibt sich die Frage, ob die beiden Koeffizientenreihenfolgen auch in Bezug auf die Aussteuerbarkeit gleichwertig sind. Mit einem vom Programm abfragbaren Schalter läßt sich beim Programm IWDIM die Reihenfolge der Koeffizienten ändern. Somit ist eine erste experimentelle Beurteilung dieser Frage möglich. Als ein Maß für die Aussteuerung dient uns das Verhältnis AI zwischen der Amplitude des anregenden Einheitsimpulses und der maximal auftretenden Zustandsgröße bei der Berechnung der ersten 1024 Werte der Impulsantwort. Dabei zeigte sich, daß bei den Koeffizienten $\alpha_G = 0$, $\alpha_S = 1$, was in der elektrischen Ersatzschaltung entsprechend Bild 2.14a der Einspeisung durch eine Spannungsquelle mit Innenwiderstand und einem Kurzschluß als Abschluß entspricht, bei allen untersuchten Koeffizientensätzen eine höhere Aussteuerung als bei der im Übertragungsverhalten äquivalenten Möglichkeit mit $\alpha_G = 1$, $\alpha_S = 0$ mit geänderter

Reihenfolge der α_i möglich ist. Letzteres entspricht der in Bild 2.14c dargestellten Ersatzschaltung mit Speisung durch eine ideale Stromquelle und Abschluß durch einen Widerstand. Im ersten Fall, den wir im weiteren mit dem Index 0,1 kennzeichnen wollen, ergab sich

$$AI_{0,1} = 0{,}56 \ldots 0{,}94,$$

während für den mit dem Index 1,0 gekennzeichneten zweiten Fall die Werte den Bereich

$$AI_{1,0} = 0{,}14 \ldots 0{,}41$$

überstreichen. Der Quotient der Aussteuerungen für den jeweils gleichen Koeffizienten ergab

$$AI_{0,1}/AI_{1,0} = 1{,}4 \ldots 5{,}7 \qquad (5.7)$$

Neben dem experimentellen Nachweis können wir auch eine theoretische Begründung für das beobachtete Verhalten liefern. Falls das Anregungssignal ein konstantes Leistungsdichtespektrum besitzt, ist nämlich der Effektivwert der Zustandsgrößen bei dem vorliegenden Leitungsmodell proportional der Quadratwurzel des maximalen charakteristischen Widerstandes eines Leitungselements (MEERKÖTTER, 1980). Wie wir im Kap. 3.2.3 gezeigt haben, gilt für den Fall $\alpha_G = 1$, $\alpha_S = 0$

$$\alpha_i = k_i \ , \quad i = 1,..,p \qquad (5.8)$$

während für $\alpha_G = 0$, $\alpha_S = 1$

$$\alpha_i = k_{p-i+1} \ , \quad i = 1,..,p \qquad (5.9)$$

gilt. Anhand des rekursiven Lösungsalgorithmus (3.39) sieht man, daß

$$k_1 = - \frac{R(1)}{R(0)} \qquad (5.10)$$

der normierte Autokorrelationskoeffizient ist, während die

$-k_i$, i = 2,..,p Näherungswerte für partielle Korrelationskoeffizienten mit Zeitverzögerungen iT sind (DURBIN, 1960). In der Literatur wurden später die k_i bzw. ähnlich definierte Koeffizienten als PARCOR Koeffizienten bezeichnet (ITAKURA & SAITO, 1972), da sie bei den schon erwähnten "Lattice Methoden" (ITAKURA & SAITO, 1971; MAKHOUL, 1977) mit Hilfe der normierten Kreuzkorrelation zwischen dem Vorwärtsprädiktionsfehler $f_m(n)$ und dem Rückwärtsprädiktionsfehler $b_m(n)$ gebildet werden können (siehe Bild 3.5). Aufgrund dieser Korrelationseigenschaften wird der Betrag der k_i mit wachsendem i im Mittel immer kleiner. Mit Hilfe von

$$R_{i+1} = R_i \frac{1-\alpha_i}{1+\alpha_i} \qquad (5.11)$$

können wir die charakteristischen Widerstände der Leitungselemente für die beiden möglichen Fälle berechnen. Da die Adaptorkoeffizienten nur das Verhältnis zweier Widerstände beschreiben, müssen wir einen Widerstandswert festlegen. In unserem Fall haben wir jeweils

$$R_1 = 1$$

gewählt. Wie wir aufgrund der experimentellen Ergebnisse über die Aussteuerung erwarten konnten, sind die $R_{max\ 1,0}$ bei den untersuchten Koeffizientensätzen immer größer als die $R_{max\ 0,1}$. Die den Quotienten der jeweils zu einem Koeffizientensatz gehörenden Aussteuerungen (5.7) entsprechenden Wurzeln des Verhältnisses der maximalen Widerstände überstreichen mit

$$\sqrt{\frac{R_{max1,0}}{R_{max0,1}}} = 1{,}8 \ \ldots \ 6{,}1 \qquad (5.12)$$

fast den gleichen Bereich und entsprechen für den jeweiligen Koeffizientensatz dem Aussteuerungsverhalten recht gut. Im allgemeinen lassen sich also auch ohne Simulation auf dem Rechner Aussagen über das Aussteuerungsverhalten des Leitungsmodells anhand der jeweiligen Koeffizientensätze bzw. der charakteristischen Widerstände der Leitungselemente machen.

5.3 Maßnahmen zur Sicherung der Stabilität

Aufgrund der digitalen Realisierung mit endlicher Wortlänge und der dadurch eingeführten Nichtlinearitäten sind die für das lineare System geltenden Stabilitätsbedingungen nicht mehr hinreichend. Im Gegensatz zu anderen digitalen Strukturen kann bei Wellendigitalfiltern, wie wir schon in Abschnitt 2.2 erwähnt haben, mit Hilfe der Pseudoleistung (FETTWEIS, 1972a) die sogenannte Pseudopassivität definiert werden. Mit ihrer Hilfe kann gezeigt werden, daß ein pseudopassives Wellendigitalfilter stabil ist, d.h., daß bei verschwindendem Eingangssignal das Ausgangssignal gegen Null konvergiert. Die weiterreichende Eigenschaft der Antwortstabilität - das Abklingen einer durch Überlauf verursachten Störung auch bei nicht sehr kleinem Eingangssignal - kann erst dann gewährleistet werden, wenn geeignete Überlaufcharakteristiken verwendet werden (CLAASEN et al., 1975; MEERKÖTTER, 1979). Diese zweite Eigenschaft erfordert daher einen etwas größeren Aufwand bei einer Realisierung mit Festkomma-Arithmetik, während zur Erzielung der Pseudopassivität sehr einfache schaltungstechnische, bzw. bei der Simulation programmtechnische Maßnahmen ausreichen (FETTWEIS & MEERKÖTTER, 1975a; FETTWEIS & MEERKÖTTER, 1976). Die Bedingung, die bei der von uns realisierten und als Rechnerprogramm simulierten Schaltung erfüllt werden muß, lautet

$$|b_\nu| \leq |b_{\nu o}| \tag{5.13}$$

Dabei sind die $b_{\nu o}$, die unter idealen linearen Bedingungen an den Toren der jeweiligen Adaptoren auftretenden reflektierten Wellen, während die b_ν die entsprechenden Größen bei der Festkommarealisierung sind. Sowohl bei der aufgebauten Schaltung als auch bei der Rechnersimulation sind die auftretenden Zahlen im Zweierkomplement dargestellt. Eine aus m bit gebildete Zahl wird also durch

$$-\delta_o + \sum_{\mu=1}^{m-1} \delta_\mu \, 2^{-\mu}$$

dargestellt, wobei δ die Werte Null oder Eins annehmen kann. Die größte darstellbare Zahl beträgt somit $1 - 2^{-m+1}$ und die kleinste -1. Bei dem aufgebauten Zweitoradaptor, dessen Blockschaltbild in Bild 2.5b dargestellt ist, wird die Einhaltung der Bedingung (5.13) durch folgende Maßnahmen erreicht:

1. Abschneiden des aus 25 bit bestehenden Multiplikationsergebnisses auf die zur Weiterverarbeitung maximal zulässiger Länge von 16 bit.

2. Verringerung des Betrages von b_1 bzw. b_2 um das niedrigstwertige bit, wenn $\delta_o = 1$, d.h. die Zahl negativ ist. (Betragsschneiden)

Um einen Eindruck davon zu gewinnen, wie sich diese Maßnahmen auswirken, haben wir in dem Simulationsprogramm die Möglichkeit geschaffen, die Berechnung der Ausgangsgrößen ohne die oben aufgeführte zweite Maßnahme durchführen zu lassen. Dadurch wird die Bedingung (5.13) bei negativen Abtastwerten der Wellen b_1 und b_2 der jeweiligen Adaptoren in den meisten Fällen nicht erfüllt. Der einzige Fall, in dem gerade das Gleichheitszeichen bei (5.13) Gültigkeit hat, liegt dann vor, wenn $\delta_\nu = 0$ für $\nu = 15$ bis 24 ist.

Bild 5.2 zeigt eine Impulsantwort h(n) des Leitungsmodells für n = 110 bis 1024, die mit Hilfe der Rechnersimulation für drei verschiedene Fälle gewonnen wurde. Dabei wurde der benutzte Koeffizientensatz willkürlich gewählt. Der maximal mögliche Betrag der Abtastwerte der Impulsantwort und der Zustandsvariablen ist Eins. In diesem Fall betrug

$$|h(n)|_{max} = 0{,}36 \qquad n = 1,\ldots,1024$$

und

$$|ZUST|_{max} = 0{,}81 \ ,$$

wobei die Amplitude des Eingangseinheitsimpulses

$$AMP = 0{,}5$$

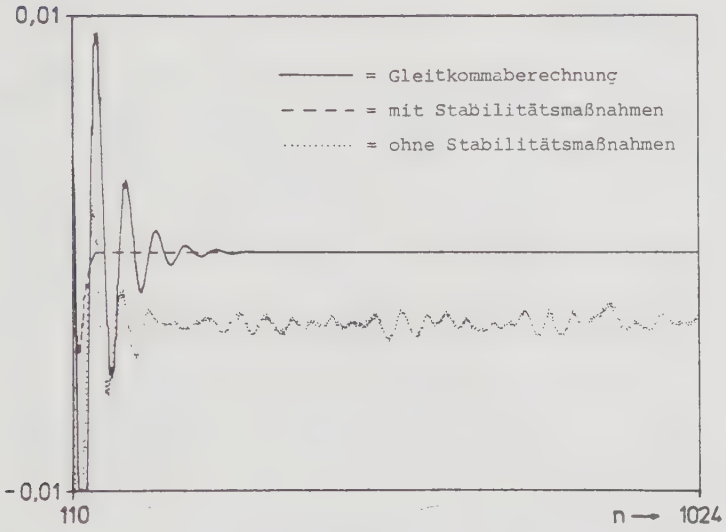


Bild 5.2 Einfluß der Maßnahmen zur Stabilitätssicherung auf die Impulsantwort des Filters

betrug. Wie wir in Bild 5.2 sehen, erreicht die Impulsantwort der Festkommarealisierung mit den beiden zur Sicherung der Stabilität nötigen Maßnahmen schon erheblich früher als die dem idealen linearen Fall recht nahe kommende Gleitkommarealisierung den Wert Null. Die Impulsantwort der Festkommarealisierung ohne die zweite oben aufgeführte Maßnahme folgt, wie wir sehen, zwar länger dem idealen linearen Fall, führt aber dann parasitäre Schwingungen um einen negativen Mittelwert aus. Dieser negative Mittelwert läßt sich dadurch erklären, daß ohne die zweite Maßnahme

$$|b_\nu| \geq |b_{\nu o}| \quad \text{für} \quad b_\nu < 0 \tag{5.15}$$

ist. Die maximale Schwankungsbreite beträgt in dem dargestellten Fall 53 Quantisierungsstufen, so daß einmal von dem Mittelwert abgesehen die 6 niedrigwertigsten bits durch die parasitäre Schwingung beeinflußt werden.

6. Erste Synthesebeispiele

Zusammen mit einem Programm, das die gewonnenen Sprachgrundfrequenzwerte mit einer frei wählbaren Anregungsfunktion zu einem Eingangssignalfile verarbeitet, bildet das schon erwähnte Programm IWDSY ein System, das das Sprachsynthesesystem simuliert. Für die Erzeugung von Synthesebeispielen mit dem schaltungsmäßig aufgebauten Leitungsmodell und Anregungsgenerator wurde ein Steuerungsprogramm erstellt. Mit ihm werden die gewonnenen Parameter für die Übergabe an die Schaltung vorbereitet und zur Spracherzeugung an sie übergeben. Mit beiden Systemen wurden erste Beispiele synthetischer Sprache erzeugt. Dabei wurden die 13 Adaptorkoeffizienten aus Analysefenstern der Länge NW = 256, die um NP = 128 Abtastwerte verschoben wurden und mit einem Hamming Fenster bewertet wurden, gewonnen. Bei der Gewinnung der Sprachgrundfrequenz wurde ein Analysefenster der Länge NW = 256 benutzt, das jeweils um die gesamte Länge verschoben wurde. Somit stehen bei der Abtastfrequenz von f_s = 10 kHz dem Filter alle 12,8 msec neue Parameter zur Verfügung, während sich das Anregungssignal nur alle 25,6 msec ändert. Diese Zahlenwerte wurden zunächst einmal auf Grund der Ergebnisse von Kapitel 3 und unter Berücksichtigung der Formate der Datenspeicherung unseres Laborrechners ausgewählt. Sie erwiesen sich aber auch bei den ersten erzeugten Synthesebeispielen als sinnvoll. Eine genauere Optimierung ist erst durch umfangreiche Hörversuche möglich.

Mit den ersten erzeugten Synthesebeispielen wurden vier Anregungsfunktionen einer ersten Beurteilung unterzogen. Sie sind in Bild 6.1 dargestellt. Dabei ist die Folge a das aus einem von KUREMATSU et al. 1979 durchgeführten Test als am häufigsten bevorzugtes Anregungssignal hervorgegangen, während die Folge b bei einer Untersuchung von SAMBUR et al. 1978 als geeignetstes Anregungssignal angesehen wurde. Die Folge c ist das von WONG & MARKEL, 1978 vorgestellte Anregungssignal, bei dem besonders auf die Erhaltung der Phase des natürlichen Anregungssignals geachtet wurde. Das vierte Anregungssignal ist der unter d dargestellte Einheitsimpuls. Bild 6.2 zeigt die entsprechenden logarithmierten Betragsspektren der vier Signale. Zur Berechnung

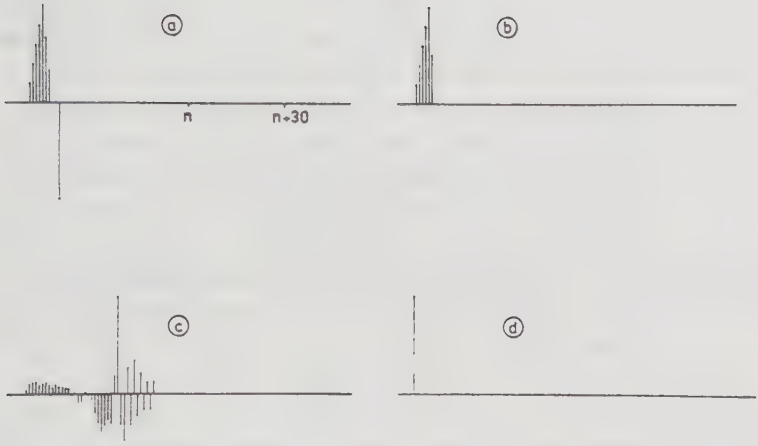


Bild 6.1 Vier zur Synthese benutzte Anregungsfolgen

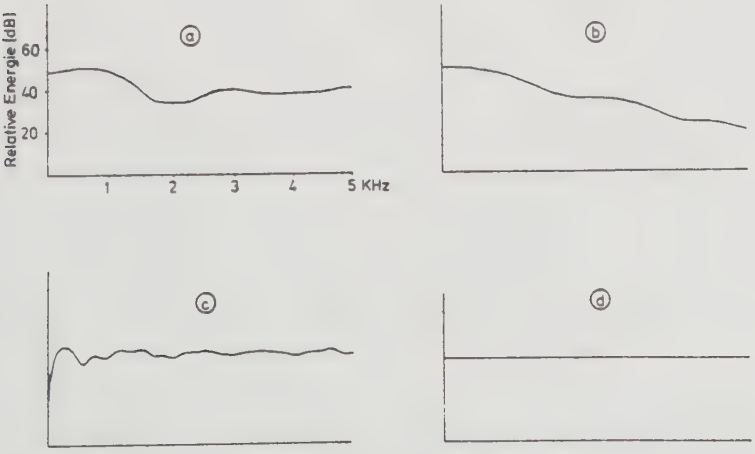


Bild 6.2 Logarithmierte Betragsspektren der in Bild 6.1 dargestellten Anregungsfolgen

der diskreten Fouriertransformation wurden die Folgen bis zum Abtastwert n = 1024 mit Nullen ergänzt. Ein erstes vorläufiges Ergebnis der Hörversuche läßt sich auch an den Spektren erkennen. Die zu hohen Frequenzen abfallenden Spektren der beiden ersten Anregungssignale machen sich bei der erzeugten Sprache durch einen unnatürlichen dumpfen Klang bemerkbar. Somit erscheinen nach diesen ersten Hörversuchen in Bezug auf die Natürlichkeit der erzeugten Sprache nur die Anregungsfunktionen c und d als geeignet. Um aus ihnen die zur Erzeugung natürlich klingender Sprache geeignetste Anregungsfunktion zu ermitteln, bedarf es umfangreicher Hörversuchsreihen mit geeignetem Sprachmaterial.

Die Natürlichkeit der ersten erzeugten Beispiele ist schon so hoch, daß der Sprecher der analysierten Sprachproben als "Sprecher" der synthetisch erzeugten Sprache erkannt wird.

7. Zusammenfassung

Die vorliegende Arbeit befaßt sich mit der Einsatzmöglichkeit eines digitalen Leitungsmodelles als Hilfsmittel zur synthetischen Spracherzeugung. Ausgehend von der elektrischen Ersatzschaltung einer akustischen Leitung im Frequenzbereich der Ausbreitung ebener Wellen, läßt sich z.B. der inhomogene Vokaltrakt durch die Aneinanderreihung gleichlanger Leitungsstücke mit jeweils konstantem Querschnitt beschreiben. Betrachtet man nur die verlustfreie Ausbreitung, so setzt sich ein solches Leitungsmodell aus Leitungselementen zusammen. Bei einer digitalen Realisierung bietet sich die Theorie der Wellendigitalfilter an, in der es ein entsprechendes Bauelement gibt. Im Abschnitt 2.2 werden einige für diese Arbeit wichtige Grundlagen dieser Theorie rekapituliert und die Vorteile gegenüber anderen digitalen Strukturen - Vermeidung von parasitären Schwingungen, Antwortstabilität des Filters - dargestellt.

Bei der Betrachtung der Übertragungseigenschaften des Leitungsmodelles in Abschnitt 2.4 konnte in dieser Arbeit gezeigt werden, daß bei Invertierung der Reihenfolge des Adaptorkoeffizientensatzes sich die betrachtete Übertragungsfunktion nicht ändert.

In einem Sprachsynthesesystem kann das schaltungsmäßig realisierte Leitungsmodell zum einen in einem physiologisch begründeten Modell als Nachbildung des Vokaltraktes eingesetzt werden. Zum anderen kann es als Prädiktionsfilter in einem Modell benutzt werden, das ohne Rücksicht auf die tatsächlichen Vorgänge bei der Spracherzeugung das Ein-Ausgangsverhalten nachbildet.

Eine kritische Betrachtung der Verfahren zur Gewinnung der Querschnittsflächenfunktion des menschlichen Vokaltraktes, die für das erste Modell benötigt wird in Abschnitt 3.1, ergab, daß zur Zeit kein verläßliches Verfahren, das ohne großen experimentellen Aufwand auskommt, vorhanden ist.

Als Prädiktionsfilter kann, wie im Abschnitt 3.2.3 gezeigt werden konnte, das Leitungsmodell die gewünschte Übertragungsfunktion auf Grund eines frei wählbaren Koeffizienten durch beliebig viele unterschiedliche Adaptorkoeffizientensätze realisieren. Bei einer geeigneten Festlegung dieses Koeffizienten ergibt sich ein einfacher rekursiver Zusammenhang mit den Prädiktionskoeffizienten eines rein rekursiven Prädiktionsfilters. Nach eingehender Analyse wurde ein geeignetes Verfahren zur Ermittlung der Adaptorkoeffizienten aus natürlicher Sprache auf dem Laborrechner implementiert.

Um vor der schaltungsmäßigen Realisierung des Leitungsmodells Aussagen über die zu wählenden Größen für verschiedene Parameter zu erhalten und zur intensiven Analyse des Modells, wurde ein Simulationsprogramm auf dem Laborrechner erstellt. Damit konnte insbesondere experimentell nachgewiesen werden, daß auf Grund der Aussteuerbarkeit des Filters trotz der Identität der Übertragungsfunktionen nur eine der beiden Koeffizientenreihenfolgen vorzuziehen ist. Für dieses Verhalten konnte auch eine auf der Theorie der Wellendigitalfilter beruhende theoretische Begründung gegeben werden.

Bei der Realisierung des Generators für die Anregungsfunktion des Sprachsynthesesystems entsprechend dem zweiten Modell wurde insbesondere bei der stimmhaften Anregung auf große Flexibilität geachtet. Durch Verwendung eines RAM-Speichers können beliebige Zeitfunktionen benutzt werden. Somit kann bei dem geplanten Einsatz in einem System zur Sprachausgabe aus Text mit unbegrenztem Wortschatz in Hörversuchen z.B. die Natürlichkeit der erzeugten Sprache in Abhängigkeit von der Anregungsfunktion optimiert werden. Die Optimierungsmöglichkeit in Hörversuchen ist bei dem gesamten schaltungsmäßig aufgebauten Sprachsynthesesystem, das vom Rechner gesteuert wird, auf Grund des flexiblen Aufbaues vorhanden. Außerdem wurde es auch als Simulation auf dem Laborrechner realisiert.

Mit den implementierten Programmen zur Gewinnung aller nötigen Parameter aus natürlicher Sprache wurden die nötigen Daten für erste Synthesebeispiele gewonnen.

Literaturverzeichnis

ATAL, B.S. (1970): Determination of the vocal-tract shape directly from the speech wave, J. Acoust. Soc. Am. 47 (A), S. 65

ATAL, B.S. (1974): Towards determining articulator positions from the Speech Signal, Proc. Speech. Comm. Seminar Stockholm, Schweden, S. 1-9

ATAL, B.S.; HANAUER, S. (1971): Speech Analysis and Synthesis by Linear Prediction of the Speech Wave, J. Acoust. Soc. Am. 50, S. 637-655

BOX, G.E.; JENKINS, G.M. (1970): Time Series Analysis Forecasting and Control, Holden-Day, San Francisco

BRAAS, J. (1977): Application of the wave digital filter concept to speech synthesis, Proc. 9th Int. Congress on Acoustics, Madrid, Juli 1977, S. 515

BURGHARDT, J. (1978): Realisierung eines rechnergesteuerten, digitalen Allpolfilters in Wellendigitalfilterstruktur als Sprachsynthesator, Diplomarbeit, Lehrstuhl für allgemeine Elektrotechnik und Akustik, Ruhr-Universität Bochum

CARCAUD, M.; COURBON, J.L.; GENIN, J.; LUCAS, J.P. (1976): A Hardware Vocal Source Simulator, Proc. 1976 IEEE Int.Conf. on Acoustics, Speech & Signal Processing, Philadelphia, Pa., S. 51-54

CHANDRA, S.; LIN, W. (1974): Experimental Comparison between Stationary and Nonstationary Formulations of Linear Prediction Applied to Voiced Speech Analysis, IEEE Transactions on Acoustics, Speech & Signal Processing, ASSP-22, S. 403-415

CLAASEN, T.A.C.M.; MECKLENBRÄUCKER, W.F.G.; PEEK, J.B.H. (1975): On the stability of the forced response of digital filters with overflow nonlinearities, IEEE Trans. Circuits and Systems CAS-22, S. 692-696

CLAASEN, T.A.C.M.; MECKLENBRÄUCKER, W.F.G.; PEEK, J.B.H. (1976): Effects of quantization and overflow in recursive digital filters, IEEE Trans. Acoustics, Speech & Signal Processing, ASSP-24, S. 517-529

COKER, C.H.; PRUZANSKY, S. (1970): Sort board for random access of auditory stimuli, J. Acoust. Soc. Am. 47, S. 95(A)

DESCOUT, R.; TOUSIGNANT, B.; LECOURS, M. (1975): Vocal Tract Area Function, Measurements: Two Time-Domain Methods, Conf. Rec. IEEE Int. Conf. on Acoustics, Speech & Signal Process., S. 75-78

DUNN, H.K. (1950): The Calculation of Vowel Resonances and an Electrical Vocal Tract, J. Acoust. Soc. Am. 22, S. 746-750

DURBIN, J. (1960): The fitting of time-series models, Rev. Inst.Int. Statist., Bd. 28, S. 233-243

FANT, G. (1960): Acoustic Theory of Speech Production, Mouton and Co, s'-Gravenhage, The Netherlands

FARNSWORTH, D.W. (1940): High-speed Motion Pictures of the Human Vocal Cords, Bell Labs. Record Bd. 18, S. 203-208

FETTWEIS, A. (1960): Filters met willekeurig gekozen dempingspolen en Tschebyschewkarakteristiek in het doorlatgebied, Tijdschr. Nederl. Radiogen., Bd. 25, S. 337-382

FETTWEIS, A. (1969): Cascade synthesis of lossless two-ports by transfer matrix factorization, veröff. in: Network Theory (Proc. 1969 NATO Acvanced Study Institute on Network Theory) herausg. von R. Boite, Gordon and Breach, New York, 1972

FETTWEIS, A. (1970a): Entwurf von Digitalfiltern in Anlehnung an Verfahren der Klassischen Netzwerktheorie, Tagungsheft der NTG-Fachtagung "Analyse und Synthese von Netzwerken", Stuttgart, Oktober 1970, S. 17-20

FETTWEIS, A. (1970b): Some properties of pulse impedance matrices, Archiv für Elektronik und Übertragungstechnik, Bd. 24, S. 506-512

FETTWEIS, A. (1971a): Digital Filter Structures Related to Classical Filter Networks, Archiv für Elektronik und Übertragungstechnik, Bd. 25, S. 79-89

FETTWEIS, A. (1971b): Wave digital filters, Proc. of the Tale Summer School on Circuit Theory, Institute of Radio Engineering and Electronics, Czechoslovak Academy of Science, Prag, September 1971, S. 11-1 bis 11-10

FETTWEIS, A. (1972a): Pseudopassivity, sensitivity and stability of wave digital filters, IEEE Trans. Circuit Theory, CT-19, S. 668-673

FETTWEIS, A. (1972b): Scattering properties of wave-digital filters, Proceeding of the Florence Seminar on Digital Filtering, September 1972, S. 1-8

FETTWEIS, A. (1973): Reciprocity, interreciprocity and transposition in wave digital filters, Int. Journal on Circuit Theory and Applications, Bd. 1, S. 323-337

FETTWEIS, A. (1976): Realizability of digital filter networks, Archiv für Elektronik und Übertragungstechnik, Bd. 30, S. 605-615

FETTWEIS, A.; MEERKÖTTER, K. (1974): Methods for preventing overflow and grannularity oscillations in arbitrary wave digital filters, Proc. Summer School on Circuit Theory, Prag, Bd. 2, September 1974, S. 395-399

FETTWEIS, A.; MEERKÖTTER, K. (1975a): Suppression of parasitic oscillations in wave digital filters, IEEE Trans. Circuits and Systems, CAS-22, S. 239-246 und S. 575

FETTWEIS, A.; MEERKÖTTER, K. (1975b): On adaptors for wave digital filter, IEEE Trans. on Acoustics, Speech & Signal Processing, ASSP-23, S. 516-525

FETTWEIS, A.; MEERKÖTTER, K. (1976): Suppression of Parasitic Oscillations in Half-Synchronic Wave Digital Filters, IEEE Transactions on Circuits and Systems, CAS-23, S. 125-126

FETTWEIS, A.; MEERKÖTTER, K. (1977): On Parasitic Oscillations in Digital Filters Under Looped Conditions, IEEE Trans. on Circuits and Systems, CAS-24, S. 475-481

FLANAGAN, J.L. (1957): Note on the design of "Terminal-Analog" speech synthesizers, J. Acoust. Soc. Am. 29, S. 306-310

FLANAGAN, J.L. (1958): Some properties of the glottal sound source, J. Speech Hear. Res. 1, S. 99-116

FLANAGAN, J.L. (1972): Speech Analysis, Synthesis and Perception, Springer, Berlin, 2. Auflage

FLANAGAN, J.L. (1976): Computers That Talk and Listen: Man-Machine Communication by Voice, Proc. IEEE, Bd. 64, S. 405-415

FLANAGAN, J.L.; LANDGRAF, L.L. (1968): Self-Oscillating Source for Vocal-Tract Synthesizers, IEEE Transactions on Audio and Electroacoustics, AU-16, S. 57-64

FLANAGAN, J.L.; CHERRY, L. (1969): Excitation of Vocal-Tract Synthesizers, J. Acoust. Soc. Am. 45, S. 764-769

FLANAGAN, J.L.; RABINER, L.R., eds. (1973): Speech Synthesis Dowden, Hutchinson & Ross, Inc. Stroudsburg, Pa.

FLANAGAN, J.L.; ISHIZAKA, K.; SHIPLEY, K.L. (1975): Synthesis of Speech From a Dynamic Model of the Vocal Cords and Vocal tract, Bell System Technical Journal 54, S. 485-506

FLANAGAN, J.L.; ISHIZAKA, K. (1976): Automatic Generation of Voiceless Excitation in a Vocal Cord-Vocal Tract Speech Synthesizer, IEEE Transactions on Acoustics, Speech & Signal Processing, ASSP-24, S. 163-170

FRAITURE, L.; NEIRYNCK, J.J. (1969): Theory of Unit-Element Filters, Revue HF, 7, S. 325-340

FRIEDERICH, P. (1979): Entwurf und Aufbau eines rechnergesteuerten Signalgenerators unter besonderer Berücksichtigung der Anwendung in der Sprachsignalsynthese, Diplomarbeit, Lehrstuhl für allgemeine Elektrotechnik und Akustik, Ruhr-Universität Bochum

GOLD, B.; RABINER, L.R. (1969): Parallel Processing techniques for estimating pitch periodes of speech in the time domain, J. Acoust. Soc. Am. 46, S. 442-448

GOLD, B.; RADER, C.M. (1969): Digital processing of signals, McGraw-Hill, New York

GOPINATH, B.; SONDHI, M.M. (1971): Inversion of the Telegraph Equation and the Synthesis of Nonuniform Lines, Proc. of the IEEE 59, S. 383-392

HAMMING, R.W. (1977): Digital Filters, Prentice Hall, Englewood Cliffs, New Jersey

HEIKE, G. (1979): Sprachsynthese auf der Basis eines Artikulationsmodells, Tagungsbericht, 4. DFG-Kolloquium Sprachverarbeitung, 1. und 2. 10.1979, Darmstadt, S. 8-10

HEINZ, J.M.; STEVENS, K.N. (1964): Derivation of area functions and acoustic spectra from cineradiographic film of speech, Quarterly Progress Report No. 74, Research Laboratory of Electronics, MIT, S. 192

HEINZ, J.M.; STEVENS, K.N. (1965): On the Relation between Lateral Cineradiographs, Area Functions, and Acoustic Spectra of Speech, Proceedings of the Fifth International Congress on Acoustics, 1965, Liege, A 44

HELMHOLTZ, H.v. (1877): Die Lehre von den Tonempfindungen, Wissenschaftliche Buchgesellschaft Darmstadt 1968; Reprografischer Nachdruck der von R. Wachsmuth herausgegebenen 6. Auflage, Braunschweig 1913

HENNEKE, J. (1977): Realisierung eines nichtrekursiven Tiefpassfilters mit linearer Phase am Digitalrechner, Studienarbeit, Lehrstuhl für allgemeine Elektrotechnik und Akustik, Ruhr-Universität Bochum

IEEE Transactions on Audio (1968): Special Issue on Digital Filters, IEEE Transactions on Audio, AU-16, Nr. 3

ISHIZAKA, K.; FLANAGAN, J.L. (1972): Synthesis of Voiced Sounds From a Two-Mass Model of the Vocal Cords, Bell System Technical Journal 51, S. 1233-1268

ISOMICHI, Y. (1974): A Vocal Tract Model and Estimation of Cross-Sectional Areas, Electronics and Communications in Japan 57 A, S. 35-41

ITAKURA, F.; SAITO, S. (1971): Digital Filtering Techniques for Speech Analysis and Synthesis, Proceedings of the 7th Int. Congr. on Acoustics, Budapest, 1971, Paper 25-C-1, S. 261-264

ITAKURA, F.; SAITO, S. (1972): On the Optimum Quantization of Feature Parameters in the Parcor Speech Synthesizer, Proc. 1972 IEEE Conf. Speech Commun. Process., New York, S. 434-437

JOSWIG, M. (1979): Ein spezielles rechnergestütztes Meßverfahren zur Messung der menschlichen Trommelfellimpedanz unter Berücksichtigung der individuellen Querschnittsfunktion des Gehörgangs, Diplomarbeit, Lehrstuhl für Allgemeine Elektrotechnik und Akustik, Ruhr-Universität Bochum

KAISER, J.F. (1966): Digital Filters, veröffentlicht in: F.F. Kuo und J.F. Kaiser, System analysis by digital computer, J. Wiley & Sons, New York, S. 218-285

KAISER, J.F. (1976): On the limit cycle problem, Proc. 1976 IEEE Int. Symp. on Circuits and Systems, München, April 1976, S. 642-645

KÖSTER, J.-P. (1973): Historische Entwicklung von Syntheseapparaten zur Erzeugung statischer und vokalartiger Signale nebst Untersuchungen zur Synthese deutscher Vokale, Helmut Buske Verlag, Hamburg

KUREMATSU, A.; ISHIGAMI, H. KITAYAMA, S.; YATO, F.; TAMURA, J. (1979): A Linear Predictive Vocoder with New Pitch Extraction and Exciting Source, Proc. 1979 IEEE Int. Conf. on Acoustics Speech & Signal Processing, Washington, 1979, S. 69-72

LEVINSON, N. (1946): The Wiener RMS (root mean square) error criterion in filter design and prediction, J. Math. Phys. 25, S. 261-278

MAKHOUL, J. (1973): Spectral Analysis of Speech by Linear Prediction, IEEE Transactions on Audio and Electroacoustics, AU-21, S. 140-148

MAKHOUL, J. (1975): Linear Prediction: A Tutorial Review, Proc. of the IEEE 63, S. 561-580

MAKHOUL, J. (1977): Stable and Efficient Lattice Methods for Linear Prediction, IEEE Transactions on Acoustics, Speech and Signal Proc., ASSP-25, S. 423-428

MAKHOUL, J.; VISWANATHAN, R.; SCHWARTZ, R.; HUGGINS, A.W.F. (1978): A Mixed-Source Model for Speech Compression and Synthesis, Proc. 1978 IEEE International Conference on Acoustics, Speech & Signal Processing, Tulsa, S. 163-166

MARKEL, J.D.; GRAY, A.H., jr. (1974): A Linear Prediction Vocoder Simulation Based upon the Autocorrelation Method, IEEE Transactions on Acoustics, Speech & Signal Processing, ASSP-22, S. 124-134

MARKEL, J.D.; GRAY, A.H., Jr. (1976): Linear Prediction of Speech, Springer Verlag, Berlin

MATHEWS, M.V.; MILLER, J.E.; DAVID, E.E., Jr. (1961): Pitch Synchronous Analysis of Voiced Sounds, J. Acoust. Soc. Am. 33, S. 179-186

MC GONEGAL, C.A.; RABINER, L.R.; ROSENBERG, A.E. (1975): A Semiautomatic Pitch Detector (SAPD), IEEE Transactions on Acoustics, Speech & Signal Processing, ASSP-23, S. 570-574

MC GONEGAL, C.; RABINER, L.; ROSENBERG, A. (1977): A Subjective Evaluation of Pitch Detection Methods Using LPC Synthesized Speech, IEEE Transactions on Acoustics, Speech & Signal Processing, ASSP-25, S. 221-229

MEERKÖTTER, K. (1977): Canonic realization of wave digital filters involving unit elements, Int. Journal of Circuit Theory and Applications, Bd. 5, S. 395-402

MEERKÖTTER, K. (1979): Beiträge zur Theorie der Wellendigitalfilter, Dissertation, Ruhr-Universität Bochum

MEERKÖTTER, K. (1980): persönliche Mitteilung

MERMELSTEIN, P. (1967): Determination of the Vocl-Tract Shape from measured Formant Frequencies, J. Acoust. Soc. Am. 41, S. 1283-1294

MEYER-EPPLER, W. (1953): Zum Erzeugungsmechanismus der Geräuschlaute, Z. Phonetik, Bd. 7, S. 196-212

MILLER, R.L. (1956): Nature of the Vocal Cord Wave, J. Acoust. Soc. Am. 31, S. 667-677

OPPENHEIM, A.V.; SCHAFER, R.W. (1975): Digital Signal Processing, Prentice Hall, Englewood Cliffs, New Jersey

RABINER, L.R.; GOLD, B. (1975): Theory and application of digital signal processing, Prenctice Hall, Englewood Cliffs, New Jersey

RABINER, L.; CHENG, M.; ROSENBERG, A.; GONEGAL, Mc C. (1976): A Comparative Performance Study of Several Pitch Detection Algorithmus, IEEE Transactions on Acoustics, Speech & Signal Processing, ASSP-24, S. 399-418

RABINER, L.R.; SCHAFER, R.W. (1976): Digital Techniques for Computer Voice Response: Implementations and Applications, Proc. of the IEEE 64, S. 416-433

RABINER, L.R.; ATAL, B.S.; SAMBUR, M.R. (1977): LPC Prediction Error-Analysis of Its Variation with the Position of the Analysis Frame, IEEE Transactions on Acoustics, Speech & Signal Processing, ASSP-25, S. 434-442

RABINER, L.R.; SCHAFER, R. (1978): Digital Processing of Speech Signals, Prentice Hall, Englewood Cliffs, New Jersey

RICHARDS, P.I. (1948): Resistor-Transmission-Line Circuits, Proc. IRE, Bd. 36, S. 217-220

ROSENBERG, A.E. (1971): Effect of Glottal Pulse Shape on the Quality of Natural Vowels, J. Acoust. Soc. Am. 49, S. 583-590

ROSS, M.J.; SHAFFER, H.L.; COHEN, A.; FREUDBERG, R.; MANLEY, H.J. (1974): Average Magnitude Difference Function Pitch Extractor, IEEE Transactions on Acoustics, Speech & Signal Processing, ASSP-22, S. 353-362

ROTHENBERG, M. (1973): A new inverse-filtering technique for deriving the glottal air flow waveform during voicing, J.Acoust. Soc. Am. 53, S. 1632-1645

ROTHENBERG, M.; CARLSON, R.; GRANSTRÖM, B.; LINDQVIST-GAUFFIN, J. (1974): A Three-Parameter Voice Source for Speech Synthesis, Proceeding of the Speech Communication Seminar, Stockholm, 1974, Bd. 2, S. 235-243

ROTHENBERG, M.; ZAHORIAN, S. (1977): Nonlinear inverse filtering technique for estimating the glottal-area waveform, J. Acoust. Soc. Am. 61, S. 1063-1071

SAMBUR, M.R.; ROSENBERG, A.E.; RABINER, L.R.; MC GONEGAL, C.A. (1978): On reduzing the buzz in LPC synthesis, J. Acoust. Soc. Am. 63, S. 918-924

SCHROEDER, M.R. (1967): Determination of the geometry of the human vocal tract by acoustic measurements, J. Acoust. Soc. Am. 41, S. 1002-1010

SCHÜSSLER, H.W. (1973): Digitale Systeme zur Signalverarbeitung Springer Verlag, Berlin

SONDHI, M.M. (1975): Measurement of the glottal waveform, J. Acoust. Soc. Am. 57, S. 228-232

SONDHI, M.M. (1977): Estimation of Vocal-tract Areas: The need for acoustical measurements, Proc. Articulatory Modeling Symposium Grenoble, 10.-12. Juli 1977, S. 78-88

SONDHI, M.M.; GOPINATH, B. (1971): Determination of vocal-tract shape from impulse response at the lips, J. Acoust. Soc. Am.49, S. 1867-1873

STEVENS, K.N.; KASOWSKI, S.; FANT, C.G.M. (1950): An Electrical Analog of the Vocal Tract, J. Acoust. Soc. Am. 25, S. 734-742

STRUBE, H.W. (1974): Bestimmung der Querschnittsfunktion des menschlichen Stimmkanals aus dem Sprachsignal, Dissertation, Göttingen

SUCKEL, G. (1979): Kritische Untersuchung der existierenden Verfahren zur Tonhöhenbestimmung und Stimmhaft/Stimmlos/Pauseerkennung. Implementierung eines geeigneten Verfahrens auf dem Rechner, Studienarbeit, Lehrstuhl für allgemeine Elektrotechnik und Akustik, Ruhr-Universität Bochum

TOUSIGNANT, B.; LEFEVRE, J.-P.; LECOURS, M (1979): Speech Synthesis from Vocal tract Area Function Acoustical Measurements, Proc. of 1979 IEEE Int. Conf. on Acoust., Speech & Signal Process., Washington D.C., S. 921-924

TRETTER, S.A. (1976): Discrete Time Signal Processing, John Wiley & Sons, New York

ULLRICH, U. (1976): Effektive Verfahren zur Ermittlung des Signalrauschverhältnisses von Digitalfiltern unterschiedlicher Struktur, Dissertation, Ruhr-Universität Bochum

WAKITA, H. (1973): Direct estimation of the vocal tract shape by inverse filtering of acoustic speech waveform, IEEE Trans. on Audio and Electroacoustics, AU-21, S. 417-427

WAKITA, H. (1979): Estimation of Vocal-Tract Shapes from Acoustical Analysis of the Speech Wave: The State of the Art IEEE Trans. on Acoustics, Speech & Signal Processing, AU-27, S. 281-285

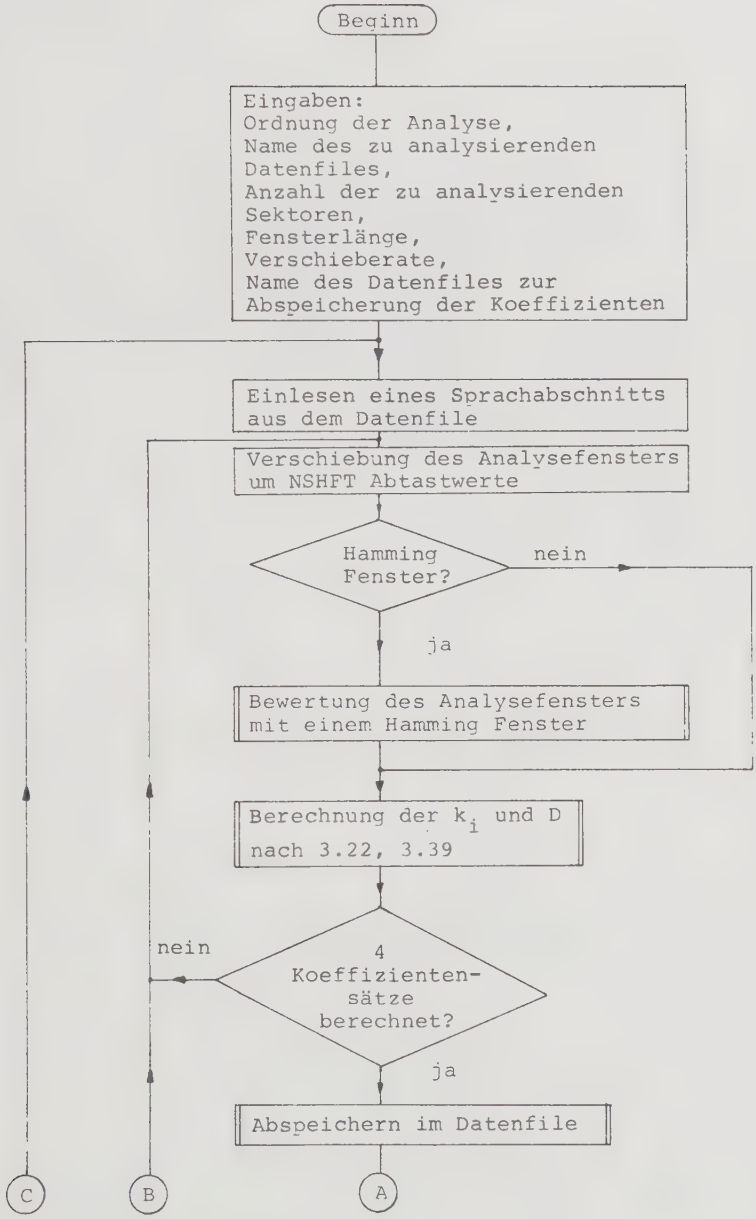
WHEATSTONE, C. (1879): The scientific papers of Sir Charles Wheatstone, Taylor and Francis, London

WIENER, N. (1949): Extrapolation, Interpolation and Smoothing of Stationary Time Series, Neuauflage mit dem Titel Time Series M.I.T. Press, Cambridge, Mass., 1966

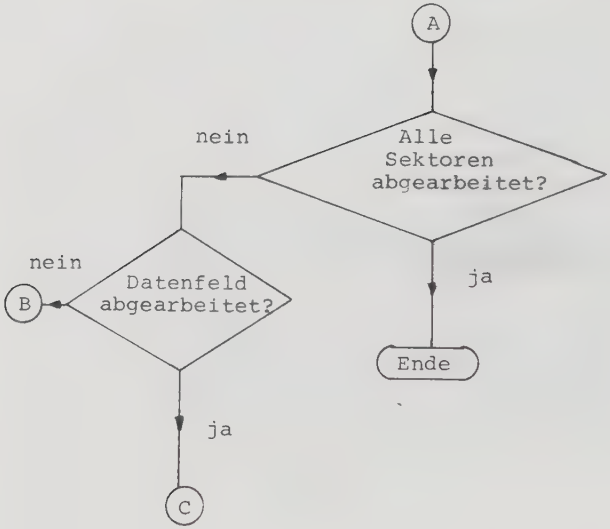
WOHLERS, M.R. (1969): Lumped and Distributed Passive Networks, Academic Press, New York

WONG, D.Y.; MARKEL, J.D. (1978): An Excitation Function for LPC Synthesis which retains the Human Glottal Phase Characteristics, Proc. 1978 IEEE Int. Conf. on Acoustics, Speech & Signal Processing, Tulsa, S. 171-174

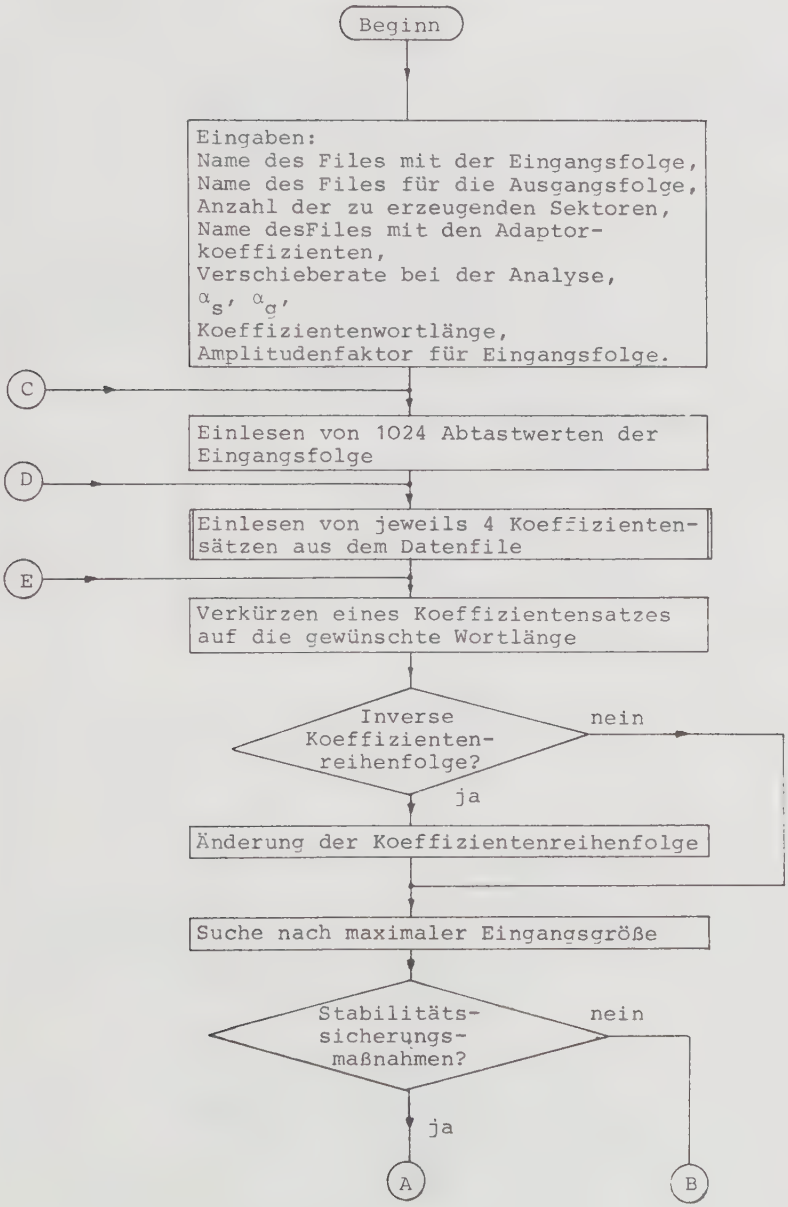
Hauptprogramm LPCAN



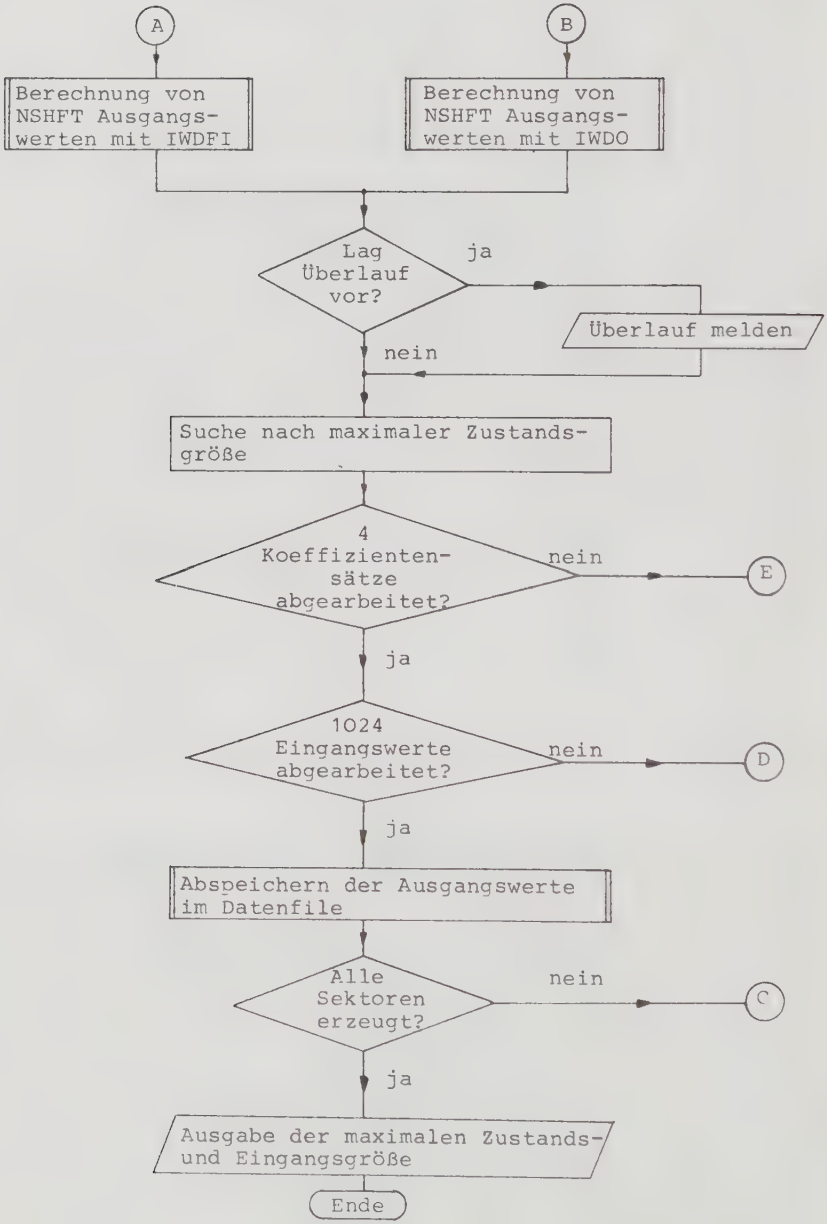
Hauptprogramm LPCAN



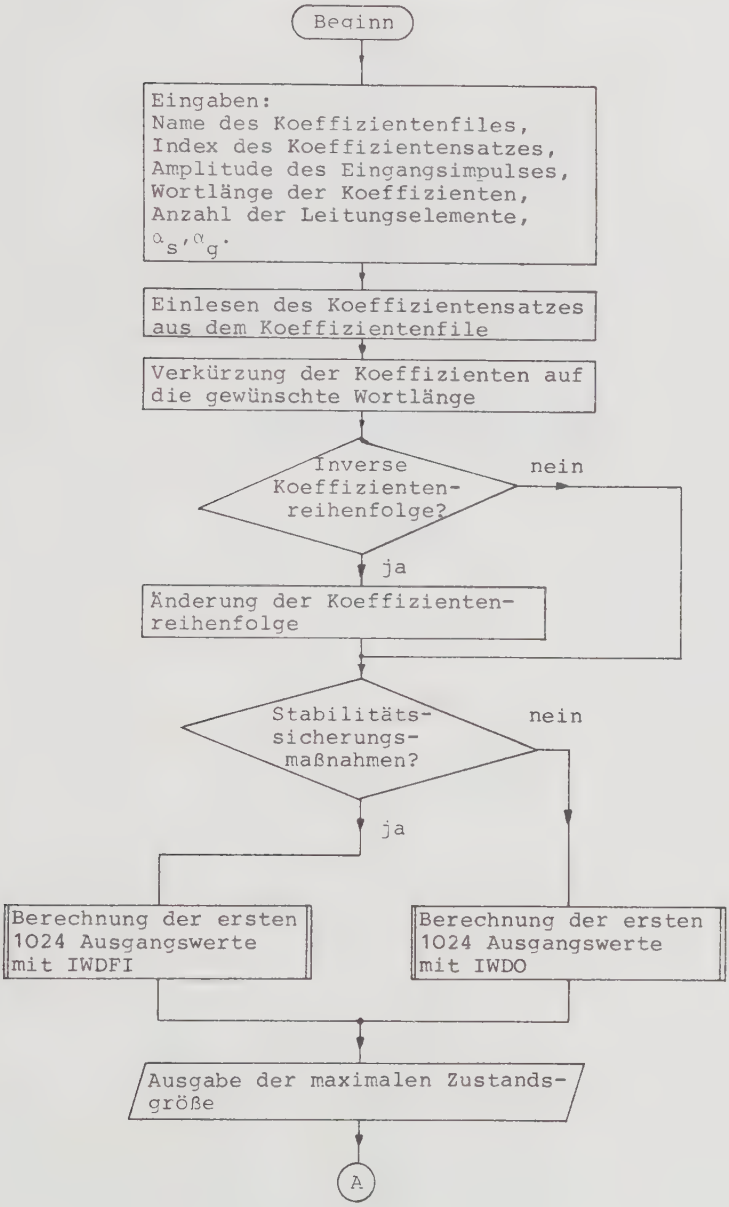
Hauptprogramm IWDSY



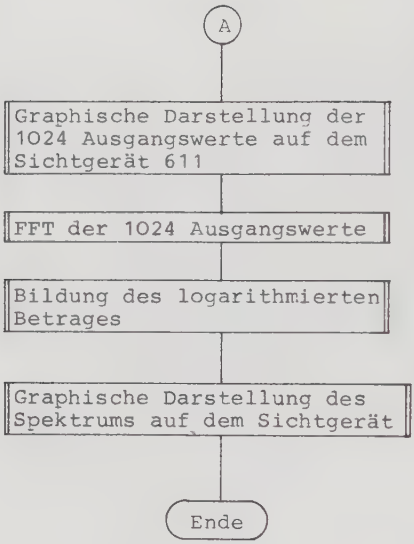
Hauptprogramm IWDSY



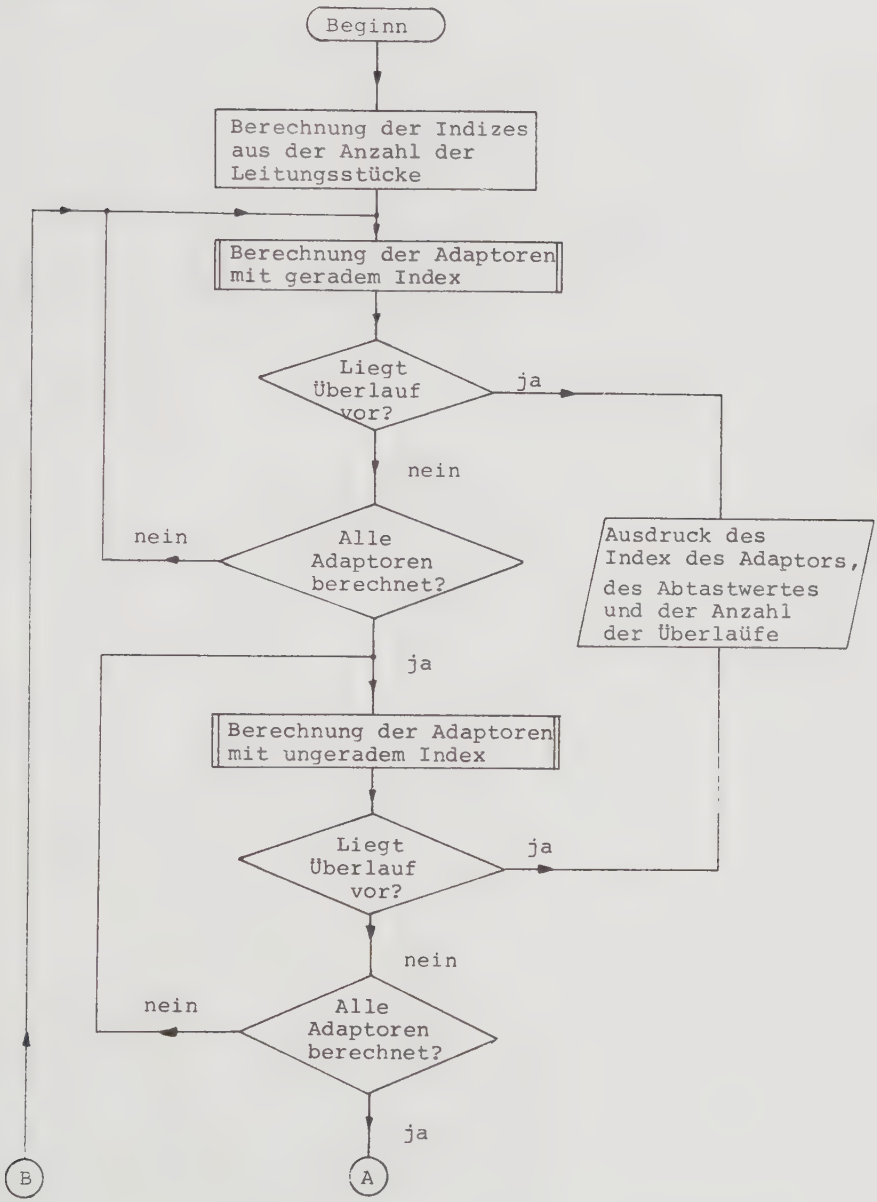
Hauptprogramm IWDIM



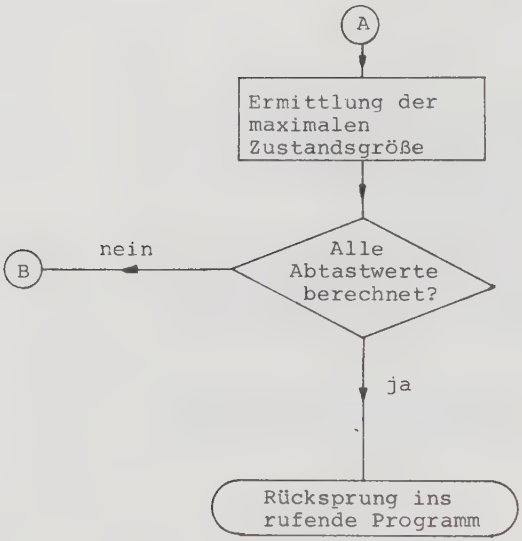
Hauptprogramm IWDIM


A
Graphische Darstellung der 1024 Ausgangswerte auf dem Sichtgerät 611
FFT der 1024 Ausgangswerte
Bildung des logarithmierten Betrages
Graphische Darstellung des Spektrums auf dem Sichtgerät
Ende

Unterprogramm IWDFI



Unterprogramm IWDFI



Lebenslauf

1948	Geboren am 02. Januar in Ahlen/Westfalen als Sohn des Ing. Helmut Braas und seiner Ehefrau Gertraud, geb. Strehlow.
1954 - 1958	Besuch der Volksschule in Oer-Erkenschwick.
1958 - 1966	Besuch des städtischen neusprachlichen Gymnasiums in Datteln. Reifeprüfung im November 1966.
1966	Industriepraktikum bei der Firma Dörnemann in Recklinghausen (November und Dezember).
1967 - 1968	Freiwilliger bei der Bundesluftwaffe.
1968 - 1973	Studium der Elektrotechnik an der Ruhr-Universität Bochum. Diplom-Hauptprüfung im November 1973.
1973 - 1980	Wissenschaftlicher Assistent zunächst am Lehrstuhl für Allgemeine und Theoretische Elektrotechnik, nach dessen Teilung am Lehrstuhl für Allgemeine Elektrotechnik und Akustik an der Ruhr-Universität Bochum.
1980	Wissenschaftlicher Mitarbeiter an einem von der DFG geförderten Projekt am Lehrstuhl für Allgemeine Elektrotechnik und Akustik.

DER DRUCKER - (Alte) Buscheystraße 138 - 4630 Bochum-Querenburg